Whole-genome sequencing of veterinary pathogens

Ronco, Troels

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

Whole-genome sequencing of veterinary pathogens

PhD thesis by Troels Ronco

December 2017

DTU Vet
National Veterinary Institute

Main supervisor

Prof. Karl Pedersen, National Veterinary Institute, Technical University of Denmark

Co-supervisor

Dr. Marc Stegger, Department of Bacteria, Parasites and Fungi, Statens Serum Institut

Assessment committee

Senior Researcher Kirstine Klitgaard Schou, National Veterinary Institute, Technical University of Denmark

Associate Prof. Peter Panduro Damborg, Department of Veterinary and Animal Sciences, University of Copenhagen

Prof. Filip Van Immerseel, Department of Pathology, Bacteriology, and Poultry diseases, University of Gent

Title of the PhD project

Whole-genome sequencing of veterinary pathogens

Submitted: 15-12-2017

This PhD Thesis is version 2 which includes the published version of Paper V

Preface

Infections in livestock caused by veterinary pathogens constitute a considerable burden in terms of decreased animal welfare and economic profit across the globe. Additionally, many veterinary pathogens possess zoonotic potential and are able to cause severe infections in humans as well. For diagnostic purposes, it is therefore important to be able to characterize these pathogenic bacteria in a fast and precise way. Specific subtypes can be identified using traditional molecular techniques based on gel-electrophoresis and single-locus sequencing. However, some of these traditional techniques do not provide the same degree of discriminatory power and reproducibility compared to whole-genome sequencing. Presently, DNA sequencing techniques have become fast and cost effective which makes them a favorable choice within the area of clinical microbiology. In this PhD project, whole-genome sequencing of three important veterinary pathogens (*Clostridium perfringens, Escherichia coli* and *Staphylococcus aureus*) was carried out and unique results that could not have been generated using traditional typing methods were obtained.

The experimental part of the project was carried out with great assistance from laboratory technicians employed at the National Veterinary Institute and Statens Serum Institut for which I am very thankful. Many sequence-analysis tools provided by Center for Genomic Epidemiology (CGE) at Technical University of Denmark were used in this project. Therefore, I will like to thank the people at CGE who have developed and maintained these valuable online-tools. This project was supported by grants from Promilleafgiftsfonden, the Danish Milk Levy Fund (Mælkeafgiftsfonden) and the Danish Poultry Levy Fond.

Furthermore, I will like to thank all scientists and academic workers who have contributed to the research papers that have been produced during this PhD project. Finally, I will dedicate a special thanks to my main supervisor Karl Pedersen and my co-supervisor Marc Stegger for excellent guidance and engagement.

Troels Ronco

Lyngby, December 2017

Contents

Summary

Bacterial infections in production animals constitute a considerable burden across the globe and result in impaired animal welfare and production economy. Additionally, veterinary pathogens with zoonotic potential may cause severe infections in humans. Traditional molecular techniques based on gel-electrophoresis and single-locus sequencing has been widely used to characterize such types of veterinary pathogens. However, DNA sequencing techniques have become fast and cost effective in recent years and whole-genome sequencing data provide a much higher discriminative power and reproducibility than any of the traditional molecular techniques. In this PhD project three important veterinary pathogens (*Clostridium perfringens, Escherichia coli* and *Staphylococcus aureus*) were investigated using whole-genome sequencing. This was done in five different scientific papers which all have been published.

Paper I and II

In 2014, an increased mortality rate caused by *Escherichia coli* infections was observed among chickens (*Gallus gallus domesticus*) from farms in Denmark and other Nordic countries. Therefore, the genetic diversity and relatedness of 114 *E. coli* isolates primary associated with increased mortality in Nordic countries, was investigated using whole-genome sequencing. In paper I, the genome assemblies of two avian pathogenic *E. coli* (APEC) isolates were annotated and the draft genomes deposited in a publicly available database. In paper II, a group of 47 closely related sequence type (ST)117 O78:H4 isolates from both broiler chickens and parent birds collected across the Nordic countries, was observed. Since most Nordic poultry farmers import birds that are all part of the same breeding pyramid these results indicate that the ST117 O78:H4 isolates were transmitted vertically through this breeding pyramide. Therefore, it was concluded that vertically transmitted ST117 O78:H4 isolates were the main reason for the increased motality rates observed in the Nordic broiler industry.

Paper III

Necrotic enteritis (NE) in chickens is primarily caused by pathogenic *Clostridium perfringens* strains. Studies have shown that three pathogenicity loci (NELoc-1, 2 and 3) and two virulence genes (*netB* and *cnaA*) are often carried by *C. perfringens* isolates collected from

chickens with NE. In general, the virulence gene content have not been widely investigated in isolates from turkeys (*Maleagris gallopavo*) with NE. However, *netB* has not been found in high prevalence among isolates from diseased turkeys and prior to this study, no publications using whole-genome sequencing on NE isolates from turkeys, have been published. Here, 30 *C. perfringens* isolates from both healthy and NE infected chickens and turkeys were analyzed using whole-genome sequencing. The results showed that NELoc-1 and -3 and the two virulence genes *netB* and *cnaA* were significantly more associated with NE isolates from chickens compared to NE isolates from turkeys. Only NELoc-2 was associated with NE isolates from both turkeys and chickens. A putative collagen adhesion gene was discovered among all isolates from diseased turkeys and has been designated *cnaD*. Potentially, CnaD could be of importance in regard to the NE pathogenesis in turkeys. In general, these results suggest that the NE pathogenesis in chickens is different from that of turkeys.

Paper IV and V

Staphylococcus aureus is a pathogen that commonly causes mastitis in dairy cows (*Bos taurus*). Many different subtypes, virulence genes and pathogenicity islands have been associated with isolates from bulk tank milk (BTM) and dairy cows with clinical mastitis (CM). Prior to these studies, no Danish *S. aureus* isolates associated with bovine mastitis have been analyzed using whole-genome sequencing. In paper V, 157 *S. aureus* isolates from BTM and dairy cows with CM were whole-genome sequenced and further investigated. In general, the results showed that BTM and CM isolates were of identical genetic background. This indicates that dairy cows can be natural carriers of *S. aureus* subtypes that in certain cases lead to CM. A group of isolates that mostly belonged to ST151 carried three pathogenicity islands that were primarily found in this group. The prevalence of resistance genes was generally low but the first ST398 methicillin resistant *S. aureus* isolate from a Danish dairy cow with CM was observed. In paper IV, the assemblies from this isolate (strain Sa52) were annotated and the draft genome uploaded to a publicly available database.

Sammendrag (Summary in Danish)

Bakterielle infektioner i produktionsdyr udgør en betydelig byrde verden over, og resulterer i både forringet dyrevelfærd og produktionsøkonomi. Derudover kan veterinære patogener med zoonotisk potentiale forårsage alvorlige infektioner hos mennesker. Traditionelle molekylære teknikker baseret på gel-elektroforese og enkelt-locus sekventering har i vid udstrækning været anvendt til at karakterisere sådanne typer af veterinære patogener. Imidlertid er DNA-sekventerings teknikkerne blevet hurtige og omkostningseffektive de seneste år, og helgenomsekventerings data tilvejebringer større diskriminerende fordele og højere reproducerbarhed sammenlignet med traditionelle molekylære teknikker. I dette ph.d. projekt blev tre vigtige veterinære patogener (*Clostridium perfringens, Escherichia coli og Staphylococcus aureus*) undersøgt ved anvendelse af helgenomsekventering. Dette blev gjort i fem forskellige videnskabelige artikler hvoraf alle er publicerede.

Artikel I og II

I 2014 blev en øget dødelighed, forårsaget af *Escherichia coli* infektioner, observeret på kyllingefarme i forskellige nordiske lande. Derfor blev den genetiske mangfoldighed samt slægtskabet i blandt 114 *E. coli* isolater, der hovedsageligt var relateret til øget dødelighed i nordiske lande, undersøgt ved anvendelse af helgenomsekventering. I artikel I blev generne i to fuglepatogene *E. coli* isolater identificeret og genomerne deponeret i en offentligt tilgængelig database. I artikel II blev der observeret en gruppe af 47 nært beslægtede ST117 O78:H4 isolater, fra både slagtekyllinger og moderfugle, indsamlet i flere nordiske lande. Da de fleste nordiske kyllingeproducenter importerer fugle, der alle er del af samme avlspyramide, indikerer disse resultater at ST117 O78:H4 isolaterne blev transmitteret vertikalt gennem denne avlspyramide. Det blev derfor konkluderet, at vertikalt overførte ST117 O78:H4 isolater var hovedårsagen til den øgede dødelighed der blev observeret i blandt nordiske fjerkræsbesætninger.

Artikel III

Nekrotiserende enteritis (NE) hos kyllinger skyldes primært patogene *Clostridium perfringens* stammer. Undersøgelser har vist, at tre patogenicitets loci (NELoc-1, 2 og 3) og to virulensgener (*netB* og *cnaA*) ofte bæres af *C. perfringens* isolater indsamlet fra kyllinger med NE. Generelt er virulensindholdet ikke blevet undersøgt dybdegående i isolater fra kalkuner

med NE. Imidlertid, er *netB* ikke blevet fundet i høj forekomst blandt isolater fra syge kalkuner, og forud for dette studie er der ikke offentliggjort publikationer med helgenomsekvenserede NE isolater fra kalkuner. Her blev 30 *C. perfringens* isolater fra både raske og NE inficerede kyllinger og kalkuner analyseret vha. helgenomsekvensering. Resultaterne viste, at NELoc-1 og -3 og de to virulensgener *netB* og *cnaA* var signifikant mere associeret med NE isolater fra kyllinger sammenlignet med NE isolater fra kalkuner. Kun NELoc-2 var relateret til NE isolater fra både kalkuner og kyllinger. Et formodet kollagen-adhæsionsgen blev opdaget i blandt alle isolater fra syge kalkuner og er blevet betegnet *cnaD*. Potentielt set kunne CnaD være af betydning ift. NE patogenesen i kalkuner. Generelt tyder disse resultater på at NE patogenesen hos kyllinger er forskellig fra den i kalkuner.

Artikel IV og V

Staphylococcus aureus er en patogen, som ofte forårsager mastitis hos malkekøer. Mange forskellige subtyper, virulensgener og patogenicitets øer er blevet associeret med isolater fra tankmælk (TM) og malkekøer med klinisk mastitis (KM). Forud for disse studier er ingen danske *S. aureus* isolater associeret med mastitis i kvæg blevet analyseret ved anvendelse af helgenomsekvensering. I artikel V blev 157 *S. aureus* isolater fra TM og mælkekøer med KM helegenomsekventeret og yderligere analyseret. Generelt viste resultaterne at TM og KM isolaterne var af identisk genetisk baggrund. Dette tyder på at mælkekøer kan være naturlige bærere af *S. aureus* subtyper, der i visse tilfælde fører til KM. En gruppe af isolater, der hovedsageligt tilhørte ST151, bar tre patogenicitets øer som primært blev fundet i denne gruppe af isolater. Udbredelsen af resistensgener var generelt lav, men det første ST398 methicillin resistente *S. aureus* isolat fra en dansk malkeko med KM blev observeret. I artikel IV blev generne i dette isolat (stamme Sa52) identificeret og genomet deponeret til en offentligt tilgængelig database.

Abbreviations

APEC: Avian pathogenic *Escherichia coli*

BLAST: Basic local alignment search tool

bp: base-pairs

BTM: Bulk tank milk

CC: Clonal complex

CM: Clinical mastitis

KM: klinisk mastitis

MALDI-TOF: Matrix-assisted laser desorption/ionization time of flight mass spectrometry

Mb: Megabases

MGE: Mobile genetic elements

MLST: Multilocus sequence typing

MRSA: Methicillin resistant *Staphylococcus aureus*

NASP: Northern Arizona SNP pipeline

NCBI: National center for biotechnology information

NE: Necrotic enteritis

NELoc: Necrotic enteritis locus

NGS: Next generation sequencing

ORF: Open reading frame

PGAAP: Prokaryotic genome automatic annotation pipeline

SaPI: *Staphylococcus aureus* pathogenicity island

SNP: Single nucleotide polymorphism

ST: Sequence type

TM: Tankmælk

1. Objectives and Papers

In this PhD project whole-genome sequencing was carried out to investigate three important veterinary pathogens (*C. perfringens*, *E. coli* and *S. aureus*). This was done in five different papers. The titles of these papers and their specific objectives can be found below whereas the original full-length papers are found in the appendixes. In the following "Background" section the main focus will therefore be on these three important veterinary pathogens and the advantages/disadvantages of using whole-genome sequencing to analyze them.

Paper I

Ronco T, Stegger M, Andersen PS, Pedersen K, Li L, Thøfner ICN, Olsen RH. 2016. Draft genome sequences of two avian pathogenic *Escherichia coli* strains of clinical importance, E44 and E51. Genome Announc. 4, e00768-16.

Objective: To whole-genome sequence and annotate two avian pathogenic *E. coli* isolates of clinical importance.

Paper II

Ronco T, Stegger M, Olsen RH, Sekse C, Nordstoga AB, Pohjanvirta T, Lilje B, Lyhs U, Andersen PS, Pedersen K. 2017. Spread of avian pathogenic *Escherichia coli* ST117 O78:H4 in Nordic broiler production. BMC Genomics 18, 1–8.

Objectives: To investigate the genetic diversity and potential relatedness of avian pathogenic *E. coli* isolates associated with increased mortality and colibacillosis in Nordic countries, using whole-genome sequencing. A further objective was to trace origin and molecular epidemiology.

Paper III

Ronco T, Stegger M, Ng KL, Lilje B, Lyhs U, Andersen PS, Pedersen K. 2017. Genome analysis of *Clostridium perfringens* isolates from healthy and necrotic enteritis infected chickens and turkeys. BMC Res. Notes 10, 1–6.

Objective: To investigate the virulence associated genomic content and genetic relationship among *C. perfringens* isolates from healthy and necrotic enteritis afflicted chickens and turkeys, using whole-genome sequencing.

Paper IV

Ronco T, Stegger M, Pedersen K. 2017. Draft genome sequence of a sequence type 398 methicillin-resistant *Staphylococcus aureus* isolate from a danish dairy cow with mastitis. Genome Announc. 5, e00492-17.

Objective: To annotate the first reported case of an ST398 methicillin-resistant *S. aureus* isolate causing mastitis in a Danish dairy cow.

Paper V

Ronco T, Klaas IC, Stegger M, Svennesen L, Astrup LB, Farre M, Pedersen K. 2018. Genomic investigation of *Staphylococcus aureus* isolates from bulk tank milk and dairy cows with clinical mastitis. Vet. Microbiol. 215, 35-42.

Objectives: To investigate the genomic content and population structure of Danish *S. aureus* isolates from bulk tank milk and clinical mastitis, using whole-genome sequencing. A further objective was to investigate possible differences between these two isolate groups and to trace the origin of strains causing clinical mastitis in dairy cows.

2. Background

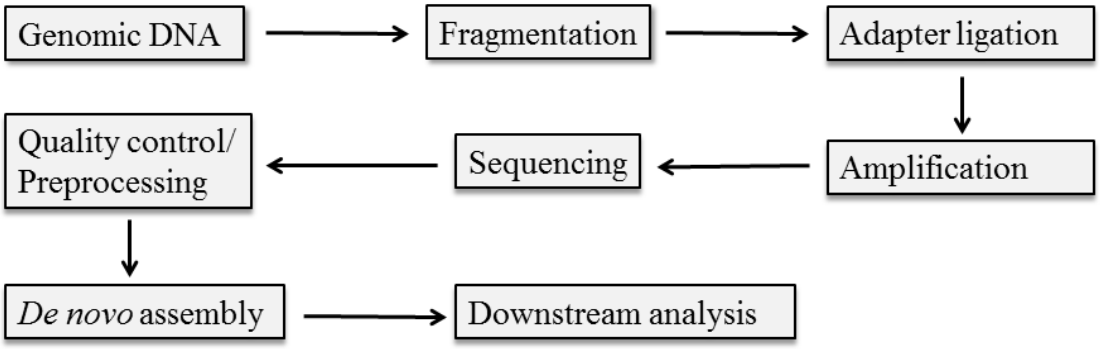
Various types of veterinary pathogens cause decreased animal welfare and economic profit on livestock farms worldwide. For example, different types of poultry infections are caused by *Clostridium perfringens* or avian pathogenic *Escherichia coli* (APEC) strains (Wade and Keyburn, 2015)(Landman and van Eck, 2015) while *Staphylococcus aureus* is known as one of the most common udder pathogens causing mastitis in dairy cows (Halasa et al., 2007). In some cases invasive pathogens with zoonotic potential such as *S. aureus* strains that belong to ST398, may spread from the livestock environment to the surrounding community where they constitute a serious threat to public health (Larsen et al., 2015). Rapid understanding and identification of such pathogens in order to diagnose and treat infections, is therefore vital for many reasons both in regard to the health of humans and livestock.

In recent years, a variety of next generation sequencing (NGS) techniques have rapidly emerged (Loman et al., 2012a). The cost and turnaround time have decreased remarkably making NGS a favorable tool within many fields of science (Didelot et al., 2012) (www.genome.gov/sequencingcosts). Specifically, within the area of clinical microbiology and epidemiology whole-genome sequencing is useful and can be applied for a wide range of purposes (Didelot et al., 2012). Concordantly, the number of bacterial genomes that are uploaded to publicly available databases hosted by the National Center for Biotechnology Information (NCBI), have increased considerably in these year (Tagini and Greub, 2017). Whole-genome sequencing data provide many pieces of information compared to analyzing just small parts of the bacterial genome amplified by traditional PCR techniques. However, sequence analysis of genomic data can be complicated and time-consuming for scientists outside the field of bioinformatics but currently, there are many user-friendly online-tools available for rapid downstream analyses of bacterial genomes (Larsen et al., 2017)(Hasman et al., 2014). Thus, microbiologists, scientists and doctors can easily analyze their own genomic data without having further experiences in bioinformatics or sequence analyses in general. Compared to data generated using some traditional molecular techniques NGS data is highly reproducible and easy to share via databases between scientists all over the globe (Jolley et al., 2004)(Maiden et al., 1998). However, whole-genome sequencing is not an error-free technique and some important drawbacks will also be elucidated in the following.

Whole-genome sequencing and gene identification

In this project, only Illumina's NextSeq and MiSeq sequencing platforms were applied. Generally, Illumina's platforms are some of the most widely used platforms for bacterial whole-genome sequencing. When chromosomal DNA has been purified there are several steps that have to be carried out before downstream analysis of the sequenced DNA can be performed. These steps will shortly be described in the following (Fig 2.1).

Figure 2.1 Basic steps in whole-genome sequencing



After purification, chromosomal DNA is fragmentized either mechanically (ultra sound) or using specific restriction enzymes. Adapters are then ligated to the DNA fragments and additionally, barcodes and terminal sequences are added to the fragments. The barcodes and terminal sequences are short nucleotide sequences and the barcodes are used to mark the DNA fragments making it possible to identify which specific sample they originate from after sequencing. The terminal sequences attach the DNA fragments to a solid phase (flow cell). While the fragments are attached to the flow cell a specific form of amplification known as solid-phase bridge amplification takes place. Here, primers bind to the adapter sequences and clusters of DNA fragments are produced on the flow cell. The DNA fragments are then sequenced and the order of the bases is determined using fluorescently labeled nucleotides that bind to the DNA fragments. During the procedure wrong bases can be generated and insertions and deletions (indels) of bases may also occur. However, the Illumina platforms provide low error rates compared to other sequencing platforms such as Roche 454 GS FLX+ and Ion Torrent. The substitution error rate has been found to be 1 per 1000 bases and < 0.001 indels per 100 bases have been observed (Loman et al., 2012a)(Loman et al., 2012b).

After DNA sequencing, raw reads are obtained in the form of fastq-files which contain information in regard to the quality of each base in the sequenced DNA. The general quality

of the fastq-files is investigated and preprocessing is performed if the quality is not acceptable. Preprocessing could for example include adapter removal and trimming (removal) of low quality bases. However, most of Illumina's sequencers remove the adapters automatically. When it is insured that the raw reads have an acceptable quality they are assembled into longer fragments, known as contigs. This is often done using a method called *de novo* assembly. Here, software programs that depend on specific mathematical algorithms are used to generate assembled contigs in form of fasta-files. Some metric parameters that are often used to evaluate the quality of a *de novo* assembly are: the average coverage, the N_{50} length and the total number of contigs. The N_{50} length of an assembly is the weighted median of the contig-lengths, whereas the average coverage is defined as the average time each position is covered by a read. The average coverage can be calculated as follows:

$$Coverage = \frac{N * L}{G}$$ N: Number of reads, L: Read Length, G: Genome/ assembly size

It should be mentioned that the assembly procedure can introduce biases and strongly depends on which specific software programs and settings that are used for the assembly. The number of genes in a *de novo* assembly can therefore vary slightly according to which software package and settings which were used. Furthermore, the length of the raw reads (read length) does also influence on the quality of the assembly since long reads are much easier to assemble compared to short ones (Loman et al., 2012a)(Earl et al., 2011). It is also possible to assemble the raw reads via mapping (comparison) to a known reference genome. However, if the reference genome is missing specific accessory genes such as antibiotic resistance or virulence genes these will not be identified using this method. After the assembly procedure various types of downstream analyses can be carries out.

During downstream analyses all types of genes can be identified in the genome which is a considerable advantage compared to having just single-locus amplicons. For example, virulence and resistance genes can be identified *in silico* which provides information in regard to the pathogenic bacterial traits (Zankari et al., 2012)(Joensen et al., 2014). However, there are some potential drawbacks associated with *in silico* identification of antibiotic resistance genes and genes in general. Not all genes are expressed which means that the genotype may not relate to the phenotype. However, in regard to resistance genes, high concordance between phenotype and genotype has been shown and it was concluded that whole-genome sequencing constitute a realistic alternative for prediction of antibiotic resistance, compared to

conventional phenotyping (Zankari et al., 2013) (Larsen et al., 2017). Furthermore, resistance can also be mediated by chromosomal mutations, for example in the DNA sequence that encodes the binding-site for the antibiotic compound. When this PhD project was conducted the online-tool Resfinder (Zankari et al., 2012) was primary used for identification of antibiotic resistance genes. At that time, Resfinder did not identify resistance caused by mutations which means that this type of acquired resistance has not been included throughout this project (Larsen et al., 2017). Currently, a new implementation in ResFinder makes it possible to identify resistance caused by mutations. ResFinder and the majority of the other CGE tools identify genes from specific databases in the query genome, using the Basic local alignment search tool (BLAST) (Altschul et al., 1997)(Larsen et al., 2017). Therefore, new genes that are associated with virulence and resistance may not be identified if they are not present in the databases, but can initially be discovered via analyses based on comparative genomics, for example (Ronco et al., 2017a).

Bacterial subtyping

Typing of specific pathogenic bacteria is important in regard to global surveillance and tracing a potential source of various types of outbreaks. It is essential to know more than just which species a specific pathogen belong to. Many traditional typing methods such as multilocus enzyme electrophoresis (MLEE), ribotyping and pulsed-field gel electrophoresis (PFGE) are based on gel-electrophoresis techniques (Maiden et al., 1998)(Aarestrup et al., 1995). PFGE was for many years the golden standard for molecular strain typing of important pathogens such as *salmonella* (https://www.cdc.gov/pulsenet/index.html) and methicillin resistant *S. aureus* (MRSA) (Murchan et al., 2003). When carrying out PFGE, bacterial DNA is cut with specific restriction enzymes and the resulting DNA fragments are separated using gel-electrophoresis. Bacterial strains can then be identified and distinguished according to the specific band pattern which is generated when DNA fragments migrate through the gel (Maiden et al., 1998)(Nauerby et al., 2003). In MLEE differences in the band pattern among DNA fragments of multiple core metabolic (housekeeping) genes is detected (Selander et al., 1986). These fragments of housekeeping genes are also called alleles. However, typing methods based on gel-electrophoresis have some fundamental disadvantages. It can be difficult to reproduce results between different laboratories and it is not easy to share the data and compare results. More importantly, the resolution is very low when looking only at the

band pattern of targets, compared to analyzing the order of base-pairs (bp) in the nucleotide sequences of alleles. Thus, multilocus sequence typing (MLST) based on nucleotide sequences of various alleles, was developed in 1998 and is currently one of the most widely applied typing methods in regard to pathogenic bacteria (Maiden et al., 1998)(Zadoks et al., 2011)(Larsen et al., 2012)(Ronco et al., 2017b)(Price et al., 2013). Publicly available databases containing several allelic profiles of bacterial species can be accessed at PubMLST (Jolley et al., 2004). The allelic profile (usually consisting of seven different alleles) for a strain indicates which specific ST it belongs to and a single nucleotide polymorphism (SNP) in one allele will result in a new ST. MLST is based on highly conserved housekeeping genes that are found in the core genome in any species, which will not be lost or exchanged in the same way as genes found in the accessory genome. Thus, MLST is very usable for evolutionary studies of population structures. However, the ST of a strain does not necessarily reveal anything about pathogenic traits such as virulence or antibiotic resistance.

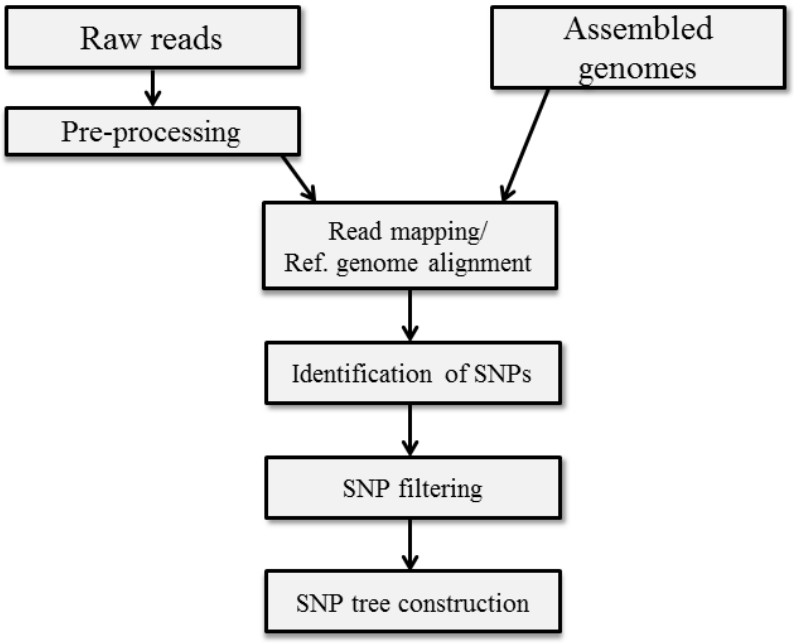
To further distinguish between different bacterial strains there are several other subtyping methods depending on the bacterial species. When investigating different types of *E. coli* strains serotyping is often used. Traditionally, this technique was based on a combination of tree different types of antigens; the flagellar antigen (H), the capsular antigen (K) and the lipopolysaccharide (O-antigen). However, K-typing is difficult to perform and therefore O:H serotyping is most often used. Generally, it is relatively difficult to perform serotyping which requires specific antiserum and a skilled laboratory technician (Joensen et al., 2015)(Roberts and Road, 1999). Another widely used typing method when looking at *S. aureus* strains is *spa*-typing. This method is based on single-locus DNA sequencing of parts (repeats) of the *spa* gene which encodes Protein A. Protein A is a surface protein involved in many aspects of the *S. aureus* pathogenesis including protection against the host immune system (Goodyear and Silverman, 2003). The repeats can be amplified from genomic DNA via traditional PCR techniques using specific primers. The amplicons can be sequenced and the nucleotide sequences uploaded to the Ridom SpaServer (spaserver.ridom.de) where it is also possible to determine the *spa*-type of bacterial isolates (Harmsen et al., 2003). Bacterial subtyping can also rely on presence of toxin encoding genes. For example, different *C. perfringens* strains are assigned a specific toxin type (A-E) according to which combinations of four major extracellular toxin genes they carry (McDonel, 1980).

Since the cost and turnaround time of DNA sequencing have decreased remarkably in recent years, whole-genome sequencing constitute an alternative to many of the above mentioned typing methods (PFGE, MLST, serotyping and *spa*-typing) which can be difficult and time consuming to carry out manually. Once the genome of a given bacteria has been sequenced, all sorts of genes depending on the bacterial species, can be extracted and related to the previously mentioned typing methods. For example, MLST, serotyping and *spa*-typing can easily be carried out *in silico* using rapid and user-friendly online-tools (Jolley et al., 2004)(Joensen et al., 2015)(Larsen et al., 2012)(Bartels et al., 2014). Furthermore, a huge advantage regarding whole-genome sequencing is that the genetic relationship between isolates can be investigated in a phylogenetic SNP analysis which can be used to distinguish further between bacteria of identical subtypes.

Phylogenetic SNP analysis

PFGE analysis has during the last decades been the golden standard for epidemiological studies and strain typing of many different pathogenic bacteria and is still widely used. However, as previously mentioned the discriminatory power of a PFGE analysis is low compared to techniques that analyze the order of bp in nucleotide sequences. Therefore, PFGE is not a suitable technique for distinguishing between closely related outbreak clones and can lead to wrong conclusions (Foley et al., 2007)(Price et al., 2013). In contrast to PFGE, phylogenetic SNP analysis is based on several variants positions in genomic regions which are conserved among all analyzed isolates. These conserved regions are also known as the core genome. Thus, this method provides high discriminatory power which makes it more suitable for distinguishing between closely related bacterial subtypes (Ronco et al., 2017b). SNP analyses can be carried out using automatic pipelines. To explain the general steps in a SNP analysis an online available pipeline called snpTree will here be explained and Figure 2.2 shows a flow diagram of the basic steps in this pipeline (Leekitcharoenphon et al., 2012).

Figure 2.2 Main steps in a snpTree-based SNP analysis pipeline



The figure shows the basic steps in a phylogenetic SNP analysis. Modified from (Leekitcharoenphon et al., 2012)

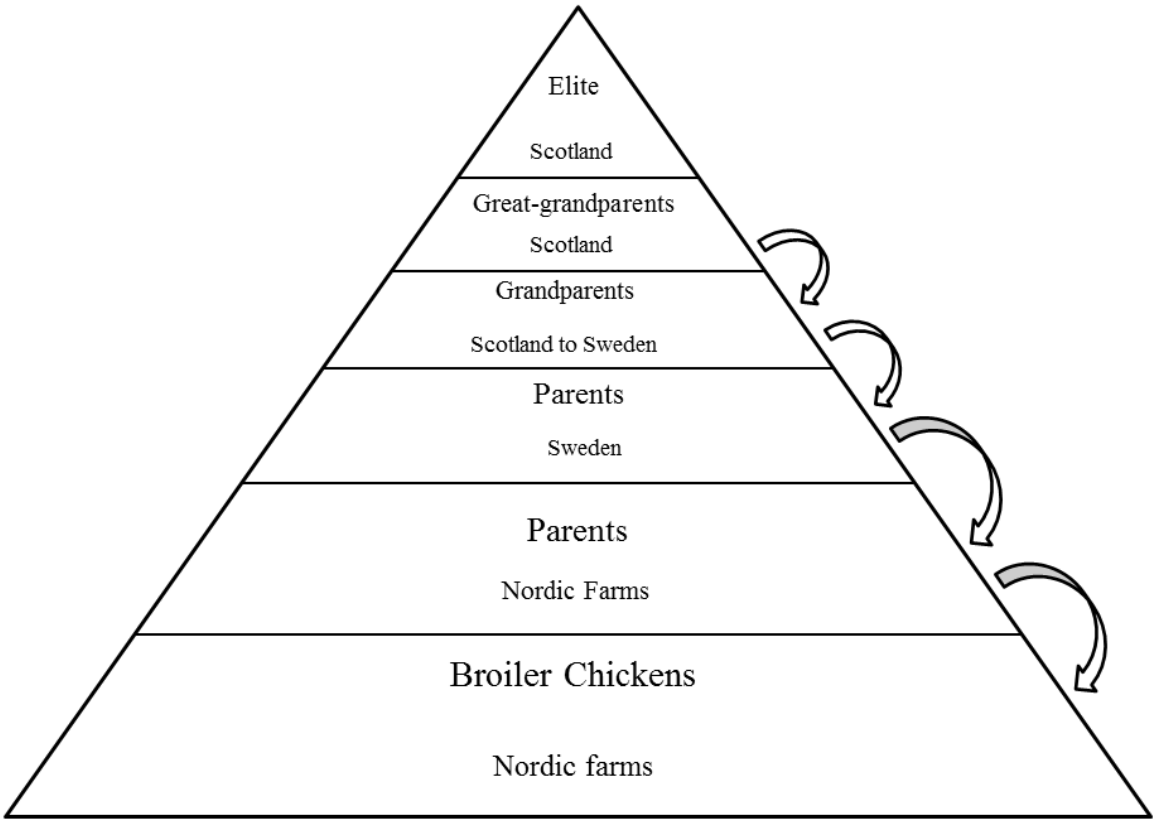
In snpTree, SNPs can be identified using both assembled genomes and raw reads. The raw reads that come directly from the sequencing machine will be preprocessed which ensures that only high quality data is used. Hereafter, the raw reads are mapped and the assembled genomes are aligned to a reference genome which is found in a database. During the mapping/aligning step core genomic regions that are present in all analyzed sequences including the reference genome are identified. Therefore, the reference genomes should preferably be relatively closely related to the analyzed isolates in order to include as much genomic material in the analysis as possible. Hereafter, variant positions are identified in these conserved genomic regions also known as the core genome. Notably, it is possible that homologue recombination occurs between large parts of two bacterial chromosomes (Stegger et al., 2013). When investigating closely related strains it can be important to exclude such recombinant chromosomal regions since they are not assumed to represent SNPs related to the core genome. Thus, SNPs that are of low quality or are assumed to reflect recombination will be excluded during a specific filtering step. Finally, a phylogenetic tree will be constructed based on the distances in SNPs between the analyzed genomes (Figure 2.2).

Avian pathogenic *Escherichia coli* and Nordic broiler production

Escherichia coli is a natural inhabitant of the intestinal tract and other mucosal layers in many different vertebrates. However, the group of extra-intestinal pathogenic *Escherichia coli* (ExPEC) causes a wide range of extra-intestinal infections in both humans and livestock. The group of ExPEC includes uropathogenic *E. coli* (UPEC), neonatal-meningitis *E. coli* (NMEC) and avian pathogenic *E. coli* (APEC) (Moulin-Schouleur et al., 2007)(Ewers et al., 2007). APEC strains cause systemic infections in poultry which is known as colibacillosis a disease that decreases the animal welfare and economic profit considerably (Landman and van Eck, 2015). Therefore, autogenous vaccines are often used to protect broilers and breeders against these types of infections (Li et al., 2017). In young broilers the disease is often seen as respiratory tract infections and fibrinous lesions in the internal organs associated with septicemia. (Dho-moulin and Fairbrother, 1999) (Ewers et al., 2007). In mature egg-laying breeders, peritonitis and salpingitis also known as egg-peritonitis is more often observed. Salpingitis is primarily caused by ascending infections with *E. coli* originating from the cloaca (Pires-dos-Santos et al., 2013) (Jordan et al., 2005). In general, APEC isolates from chickens constitute a genetically diverse population where many STs and serogroups are represented. However, it has been reported that STs 48, 95 and 117 and the serogroups O1, O2 and O78 are commonly observed among APEC strains (Dissanayake et al., 2014)(Mora et al., 2012)(Olsen et al., 2011) (Dozois et al., 1992)(Adiri et al., 2003). Interestingly, it has been reported that human ExPEC strains can be closely related to APEC strains. This means that poultry suffering from colibacillosis potentially constitute a reservoir of zoonotic APEC strains (Ewers et al., 2007).

In the Nordic countries, the majority of the broiler breeders are imported as day-old chickens from Sweden and used for further production of broilers. Therefore, Nordic chickens are part of a large broiler breeding pyramid (Mo et al., 2016)(Agersø et al., 2014) (Figure 2.2).

Figure 2.2 Broiler breeding pyramid



The figure shows that broiler chickens on most Nordic farms are part of a breeding pyramid. In the top of the pyramid are few elite birds which are used to breed great-grandparents birds. From the great-grandparents, grandparents are produced and exported from Scotland to Sweden. On Swedish hatcheries parent birds are produced and exported to farmers in the Nordic counties where broiler chickens are produced.

In this pyramid, few elite birds from Scotland are used for breeding of great-grandparents. From these great-grandparents, grandparents are produced and exported to Swedish hatcheries. In Sweden, the grandparents are then used for breeding of parent birds that are exported as day-old chickens to various Nordic broiler farms (Mo et al., 2016)(Agersø et al., 2014). Previously, studies indicate that vertical transmission of *E. coli* strains from parents to offspring can occur in such a breeding system (Zurfluh et al., 2014)(Nilsson et al., 2014).

On Danish poultry farms the mortality rate has been decreasing from 2009 until late 2014 but has hereafter started to increase remarkably. On average, the mortality increased from < 3.5% in 2014 to > 4.0% in 2015. Presumably, collibacillosis in both parent birds and broilers was the main reason (Bisgaard et al., 2015). At the same time, similar issues with increased mortality and colibacillosis cases were observed on poultry farms in other Nordic countries (Magne Hansen, Animalia, pers. comm). Therefore, the genetic background of the *E. coli*

isolates from diseased broilers and parents collected on various Nordic poultry farms were analyzed in Paper I and II.

Clostridium perfringens and necrotic enteritis in poultry

Clostridium perfringens is an anaerobic gram-positive pathogen that causes gastro intestinal diseases in both humans and livestock (Songer, 1996). In general, the virulence is due to four major toxins (α, β, ε, ι) and *C. perfringens* strains are assigned a specific toxin type (A to E) according to which of these toxins they produce (McDonel, 1980). Gastro intestinal infections caused by pathogenic *C. perfringens* strains belonging to type A and to a lesser degree type C, have been reported to be the main cause of necrotic enteritis (NE) in broiler chickens (Songer and Meer, 1996)(Nauerby et al., 2003). NE is seen as an acute clinical or a subclinical form, which causes increased mortality or decreased weight gain, respectively (Songer, 1996) (Lovland and Kaldhusdal, 2001). Altogether, this disease has a considerable negative impact on the animal welfare and production yield in the global poultry industry (Wade and Keyburn, 2015).

Studies have shown that pathogenic type A strains associated with NE in chickens express the pore-forming toxin NetB (Keyburn et al., 2010)(Keyburn et al., 2008). The intestinal microbiota of diseased chickens is dominated by one or few types of virulent *C. perfringens* strains, which is known as the single strain dominance phenomena. In contrast, the composition of *C. perfringens* strains in healthy poultry is more diverse (Nauerby et al., 2003) (Barbara et al., 2008). The *netB* gene is located on a plasmid-encoded and NE-associated pathogenicity locus named, NE locus-1 (NELoc-1) (Lepp et al., 2010). Two other loci, NELoc-2 and NELoc-3 which are found on the chromosome and on plasmids respectively, have also been found to be associated with NE in chickens. A recent study showed that a collagen adhesion gene *cnaA* located on a chromosomal locus named VR-10B, was also associated with isolates from chickens with NE (Wade et al., 2015). The expression of virulence genes located on both conjugative plasmids and chromosomes, is primarily regulated by the VirSR two-component system on the transcriptional level (Cheung et al., 2010) (Cheung et al., 2010)(Ohtani et al., 2003)(Okumura et al., 2008). The activation of the VirSR system is mediated through quorum sensing. Here, signal-peptides encoded by *agrBD* genes initiate auto-phosphorylation of the VirS sensor kinase and subsequently phosphor-

molecules are transferred to the VirR response regulator (Vidal et al., 2012)(Cheung et al., 2009).

Besides, being a major issue among broiler chickens NE is also a considerable burden in the turkey industry. Not many studies have elucidated the virulence gene content of strains from turkeys with NE (Saita et al., 2010). However, NetB does not seem to be essential for the NE pathogenesis in turkeys since studies have shown that *netB* is not present in high prevalence among isolates from diseased turkeys (Giovanardi et al., 2016)(Lyhs et al., 2013). Prior to paper III in this project, no *C. perfringens* isolates from turkeys with NE across the world have been analyzed using whole-genome sequencing.

Staphylococcus aureus and bovine mastitis

Staphylococcus aureus is an opportunistic pathogen that causes severe infections in both humans and various production animals. It is a major cause of udder infections in dairy cows which is known as bovine mastitis (Zadoks et al., 2011)(Agersø et al., 2012). Mastitis can be seen as a subclinical form which is a mild type of infection where symptoms can be difficult to recognize. In contrast, severe mammary gland infection is seen as clinical mastitis (CM). Altogether, both types of bovine mastitis cause reduced animal welfare, milk yield and economic profit on dairy farms worldwide (Halasa et al., 2007). Many STs (97, 126 133, 151, 479 and 771) (Holmes and Zadoks, 2011)(Zadoks et al., 2011) and *spa*-types (t518, t519, t524 t528, t529 and t543) have been observed among isolates from CM and bulk tank milk (BTM) (Hasman et al., 2010)(Ikawaty et al., 2009)(Sakwinska et al., 2011). A *S. aureus* strain that often has been associated with clinical mastitis is RF122 which belong to ST151 (Herron-olson et al., 2007). Strain RF122 carries various mobile genetic elements (MGSs) called *S. aureus* pathogenicity islands (SaPIs). The SaPIs vary considerably in size and carry different types of virulence genes that are considered to be involved in host specialization and the mastitis pathogenesis (Herron-olson et al., 2007). In CM and BTM isolates various types of virulence genes have been identified. These virulence genes have been reported to be associated with toxin production (*etA/B, hla/b/g, lukD/E/FS, sea-j* and *tst*), host colonization (*cap, clfA/B, cna, fib* and *sak*) and biofilm formation (*icaD* and *fnbB*). Some virulence genes found in CM isolates can also be harmful to humans. Leukocidins (encoded by *lukD/E/FS*) are involved in various types of clinical human infections, staphylococcal enterotoxins (encoded by *se* genes) can cause food poisoning and the toxic shock syndrome toxin-1

(encoded by *tst*) causes toxic shock syndrome (Lina et al., 1999)(Deurenberg et al., 2005)(Asao et al., 2003)(Umeda et al., 2017).

Furthermore, livestock associated methicillin resistant *S. aureus* (MRSA) strains belonging to ST398 have been detected in herds of dairy cows worldwide whereas this lineage has not spread among Danish cattle (Zadoks et al., 2011)(Holmes and Zadoks, 2011). However, ST398 MRSA isolates have been detected in high prevalence among Danish slaughter pigs and are now also increasingly infecting humans (Larsen et al., 2015)(Agersø et al., 2012). Previously, Danish *S. aureus* isolates have been detected in BTM using real-time PCR techniques (Katholm et al., 2012)(Mahmmod et al., 2017). Other studies have investigated the presence of genes that encode various types of exotoxins (enterotoxins, toxic shock toxins and exfoliative toxins) using conventional PCR (Larsen et al., 2000)(Larsen et al., 2002). Furthermore, Danish *S. aureus* subtypes among isolates from mastitis has been determined using various traditional typing methods but none of these methods included whole-genome sequencing (Aarestrup et al., 1995a)(Larsen et al., 2000). Prior to paper IV and V, no Danish *S. aureus* isolates from BTM or CM had been whole-genome sequenced and further analyzed in regard to genetic relationship, presence of STs and the general genomic content.

3. Methods

Bacterial isolates

In paper II, a total of 114 *E. coli* isolates collected from 88 different commercial broiler chicken farms were analyzed. If the isolates were from the same farm they were in general collected from different houses. Generally, the isolates originated from diseased broiler chickens and parents diagnosed with *E. coli* infections and only 15 isolates were from healthy birds. In 2016, 107 bacterial isolates from Danish (n = 74), Finnish (n = 15), Norwegian (n = 16) and Polish (n = 2) farms were collected and sent to the Danish Veterinary Institute for further analysis. Furthermore, genome sequences from seven isolates collected from Danish birds were kindly provided by the Danish poultry industry. In general, the samples were primarily obtained from liver and bone marrow. In paper I, the two draft genomes presented were from two Danish APEC isolates collected from diseased birds as described in paper II. These two isolates had previously been included in a Danish autogenous vaccine program.

In paper III, *C. perfringens* isolates (n = 30) from both healthy and NE infected chickens and turkeys were obtained from two previous independent studies. The isolates from healthy (n = 4) and diseased (n = 13) chickens were collected on 14 Danish farms between 1997 and 2002 by Nauerby et al., 2003 (Nauerby et al., 2003) whereas isolates from healthy (n = 4) and diseased (n = 9) turkeys were collected on seven Finnish farms between 1997 and 2010 by Lyhs et al., 2013 (Lyhs et al., 2013). The isolates were primarily collected from the intestine, but six chicken isolates originated from liver samples.

In paper V, 157 *S. aureus* isolates from BTM and dairy cows with CM were analyzed. The CM isolates (n = 63) were collected on 24 farms whereas the BTM isolates (n = 94) were collected on 91 farms. The isolates were sampled in 2016 from different Danish farms distributed across the country. The CM isolates originated from aseptic foremilk samples collected from dairy cows with CM according to the National Mastitis Council's guidelines. Subsequently, samples were sent to the Danish Veterinary Institute for *S. aureus* verification using Matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF). The BTM samples were collected as previously described and analyzed using real-time PCR test (Katholm et al., 2012). Based on the PCR test results, 100 herds with the lowest Ct-value (ranging from 21-27) were selected and samples sent to the Danish Veterinary Institute. Hereafter, samples were cultured on blood agar and random colonies suspected for being *S.*

aureus, were sub-cultured and verified as *S. aureus* using MALDI-TOF. The draft genome presented in paper IV was obtained from a CM isolate collected as described in paper V.

Sequencing, preprocessing and assembly

In all papers, Illumina's sequencing platforms were used to obtain paired-end whole-genome sequences. In paper I, II and III the MiSeq platform was used with an average read length of 2 × 251bp whereas the NextSeq platform with an average read length of 2 × 151bp, was applied in Paper IV and V. In all papers, the quality of raw reads was investigated using the FastQC software package and in case of low quality trimming was performed in CLC bio's Genomics Workbench (GW). Subsequently, *de novo* assembly was carried out using CLC bio's GW on default settings with minimum contig length of 500 nucleotides

Annotation and identification of genomic content

To predict genes present in the *de novo* assembled contigs NCBI's Prokaryotic genome automatic annotation pipeline (PGAAP) (Angiuoli et al., 2008) was used in paper I and IV. In paper III, the gene content of all isolates was investigated. Prokka (Seemann, 2014) was used to annotate genes in the assemblies and subsequently, Roary (Page et al., 2015) was used to investigate differences in presence of genes among the different isolate groups.

In all studies, various online-tools such as VirulenceFinder (Joensen et al., 2014), ResFinder (Zankari et al., 2012) and MyDBFinder (https://cge.cbs.dtu.dk/services/MyDbFinder/) were used to identify the genomic content in *de novo* assembled contigs. In other cases specific genes were extracted from annotated genomes available in NCBI's databases or obtained from other publicly available databases (Chen et al., 2016). Subsequently, these specific genes and in general genes located on > one contig, were identified in the assemblies using the BLASTN implementation in CLC bio's GW (Altschul et al., 1997).

When identifying specific genes in assemblies the length and nucleotide sequence of these genes are rarely 100% conserved in the assemblies which for example could be due to; assembly/sequencing errors and SNPs. Therefore, specific thresholds for nucleotide identity and coverage of query sequence length were used. Generally, in all papers, the open reading frames (ORFs) of the various genome collections were determined with thresholds of 90% nucleotide identity and 90% coverage of query sequence length. However, in paper V certain ORFs associated with the analyzed SaPIs, were identified with thresholds of > 80%

nucleotide identity and 90% coverage of the query sequence length. In some cases the ORFs were identified with thresholds of > 70% nucleotide identity and a SaPIs was only considered present in a genome if ≥ 80% of all its ORFs were present.

Subtyping

Isolates from all papers were subtyped *in silico* using PubMLST and MLST v1.8 (Jolley et al., 2004)(Larsen et al., 2012). Additionally, isolates from paper I and II were serotyped using SerotypeFinder (Joensen et al., 2015) and in paper IV and V isolates were *spa*-typed using spaTyper (Bartels et al., 2014). In paper III, the toxin type of the isolates was determined manually via identification of toxin encoding genes.

Statistical analyses

To analyze the differences in the virulence gene content between various isolate groups, statistical analyses were carried out. A Chi-square test for independence was generally used but in cases of ≤ 5 observations a Fisher's exact test was carried out. The confidence interval was 95% and the differences were considered significant when $P < 0.05$.

Phylogenetic analyses

In paper II, III and V the genetic relationship between the analyzed isolates were investigated using SNPs which in general were called from the raw reads. In paper II, SNPs were identified using NASP (The Northern Arizona SNP Pipeline) and *E. coli* strain CFT073 was used as reference chromosome (Roe et al., 2016). The filter step in NASP removed variant positions with < 90% unambiguous base calls and SNPs in positions that did not had a minimum depth of ≥ 10 fold. A phylogenetic tree was constructed using the maximum-likelihood algorithm in PhyML (Letunic and Bork, 2011). To further investigate the relatedness of ST117 isolates (n = 62) from paper II and ST117 isolates from NCBI's Short Read Archive (n = 21), SNPs were called as previous described but purging of recombinant regions was performed using Gubbins software package (Croucher et al., 2015).

In Paper III and V, the automatic pipeline, CSI Phylogeny (Kaas et al., 2014) was used for identification of SNPs. In paper III, *C. perfringens* strain ATCC 13124 was used as reference chromosome whereas *S. aureus* strain ED133 was used in paper V. CSI Phylogeny identified SNPs with a quality of ≥ 30, a minimum depth of ≥ 10 fold and a distance between SNPs of ≥

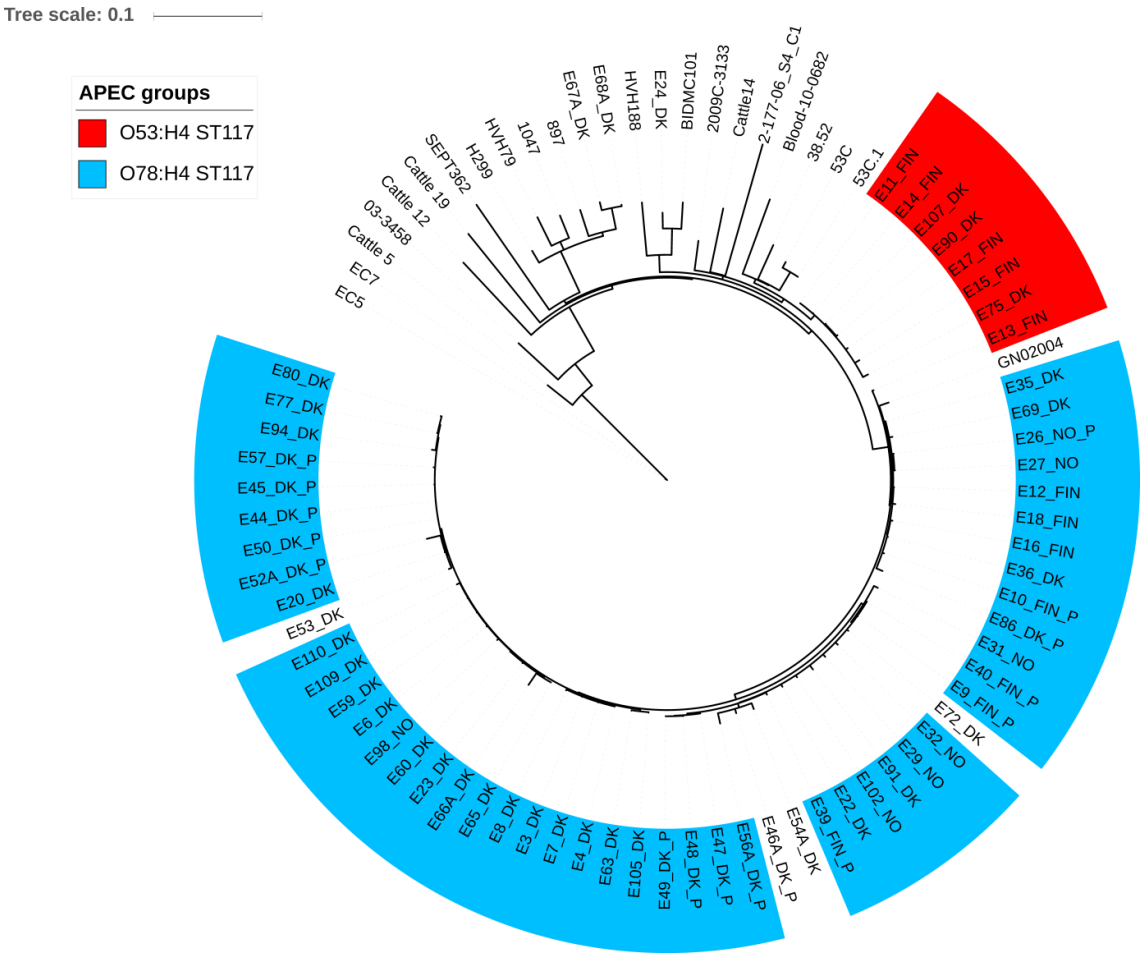
10. Subsequently, a maximum likelihood approximation was used for construction of a phylogenetic tree. In all papers, the phylogenetic trees were visualized using iTOL (Letunic and Bork, 2011).

4. Results and Discussion

Paper I and II

In paper II, the 114 *E. coli* isolates primarily associated with increased mortality and colibacillosis on Nordic chicken farms, were whole-genome sequenced and their genetic relatedness, content and diversity investigated. Fifty-eight % of the isolates were sequenced to an average coverage of ≥ 50 fold, whereas 31% had an average coverage of > 30 fold. The remaining 11% of the isolates had an average coverage of > 18 fold. In general, the results showed that the isolates constituted a genetic diverse population which correspond to previous studies (Dissanayake et al., 2014)(Olsen et al., 2011). However, a phylogenetic analysis showed a group of 47 closely related ST117 O78:H4 isolates collected from diseased broilers and parents in various Nordic countries. On average, the distance between these isolates was 23 SNPs (Figure 4.1).

Figure 4.1 Phylogenetic analysis of 83 ST117 *E. coli* isolates



The Maximum-likelihood tree includes 62 ST117 APEC isolates from this study and 21 international ST117 strains from NCBI. Forty-seven closely related O78:H4 isolates (blue) collected from both broilers and parents (P) on Danish (DK), Finnish (FIN) and Norwegian (NO) chicken farms is shown. All 62 isolates from this study were from diseased birds and the scale indicates substitutions per site.

In concordance, previous studies have shown that APEC isolates are often associated with serogroup O78 and ST117 (Olsen et al., 2011)(Ewers et al., 2007). Additionally, all ST117 O78:H4 isolates carried a unique combination of virulence genes compared to all other isolates in this study (Table 4.1).

Table 4.1 Virulence genes identified among 114 *E. coli* isolates

	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Healthy	Diseased
Isolates	47	8	5	4	3	3	15	99
Farms	34	8	4	4	2	3	15	73
Serotype	O78:H4	O53:H4	O103:H2	O18ac:H7	O149:H23	O5:H10	-	-
ST	117	117	1146	95	1163	93	-	-
fimA	+	+	+	+	+	1/3	12/15	94/99
fimC	+	+	+	+	+	+	13/15	96/99
papC								4/99
tsh				+		+		20/99
fyuA		+		3/4			1/15	28/99
iroD	+	+		+	+	+	4/15	94/99
iroN	+	+		+	+	+	3/15	93/99
irp2		+		+			1/15	30/99
iucA	+	+	+	+		+	7/15	84/99
iucD	+	+	+	+		1/3	7/15	84/99
cva		+	+	+	+	+	12/15	39/99
cvi		+	+	+	+	+	12/15	39/99
iss	+	+	2/5	+	+	+	7/15	97/99
ompA	+	+	+	+	+	+	15/15	99/99
vat	+	+		+				71/99
hlyA								
ibeA				+		1/3		13/99

The table shows different virulence genes identified among groups of isolates belonging to the same ST and serotype. Presence of genes is indicated with a "+". The number of isolates found in each group is shown together with the number of farms from where the isolates were collected. The ratios of all 114 isolates from healthy and diseased poultry that carried virulence genes are also shown. For virulence genes that were not present among all isolates in a group, the ratio of isolates that carried the genes is shown.

The parent flocks of almost all broilers in the Nordic countries originate from Swedish hatcheries where grandparents are imported from Scotland (Mo et al., 2016)(Agersø et al., 2014). Since the Nordic broiler farmers receive parent animals from the same source these findings strongly suggest vertical transmission of the ST117 O78:H4 isolates from grandparents and great grandparents in this broiler breeding pyramid. In this study, *E. coli* isolates from diseased grandparents or great grandparents in the upper part of the breeding pyramid were not available to further confirm this hypothesis. However, vertical transmission

of pathogenic *E. coli* strains has previously been observed (Agersø et al., 2014)(Zurfluh et al., 2014). There is no exact definition of a clone in regard to the distance in number of SNPs between the clones compared to the distance between isolates outside the group of clones. Since the 47 ST117 O78:H4 isolates constituted the largest group of closely related isolates with identical virulence profile and were collected from broilers and parents in multiple Nordic countries, they could here be considered as the outbreak clone.

Interestingly, in an *in vivo* study a flock of parent birds was challenged with a ST117 O78:H4 isolate from paper I and II (strain E44). This flock had a significantly lower survival rate compared to birds challenged with a different APEC strain or birds that were autogenously vaccinated. Post mortem investigations of birds from this flock showed a significantly higher pathology score regarding lesions in the salpinx, compared to all other flocks (Li et al., 2017).

In paper I, *de novo* assemblies belonging to two APEC strains (E44 and E51) that previously had been included in a Danish autogenous *E. coli* vaccine program, were annotated and added to the NCBI database. Strain E44 and E51 were also part of paper II and were sequenced to an average coverage of 91 and 58 fold respectively. The draft genome of strain E44 consisted of 195 contigs and had an assembly size of 5.1 Mb (megabases) whereas strain E51 had 217 contigs and an assembly size of 5.2 Mb. The serotype/STs of E44 and E51 were determined to be O78:H4/ST117 and O2:H5/ST140, respectively but none of the strains carried any antibiotic resistance genes.

Paper III

In paper III, *C. perfringens* isolates from healthy and NE afflicted chickens and turkeys were whole-genome sequenced and further analyzed in regard to virulence gene content and genetic relatedness. Based on a literature review it is probably the first time such types of turkey isolates have been whole-genome sequenced and made publicly available. All isolates were sequenced to an average coverage of > 40 fold except a single isolate that had > 30 fold. The results showed that two pathogenicity loci NELoc-1 and -3 and two virulence genes *netB* and *cnaA* were significantly associated with NE isolates from chickens compared to isolates from diseased turkeys. Only NELoc-2 was associated with NE isolates from both turkeys and chickens (Table 4.2).

Table 4.2 Virulence genes and loci identified in 30 *C. perfringens* strains

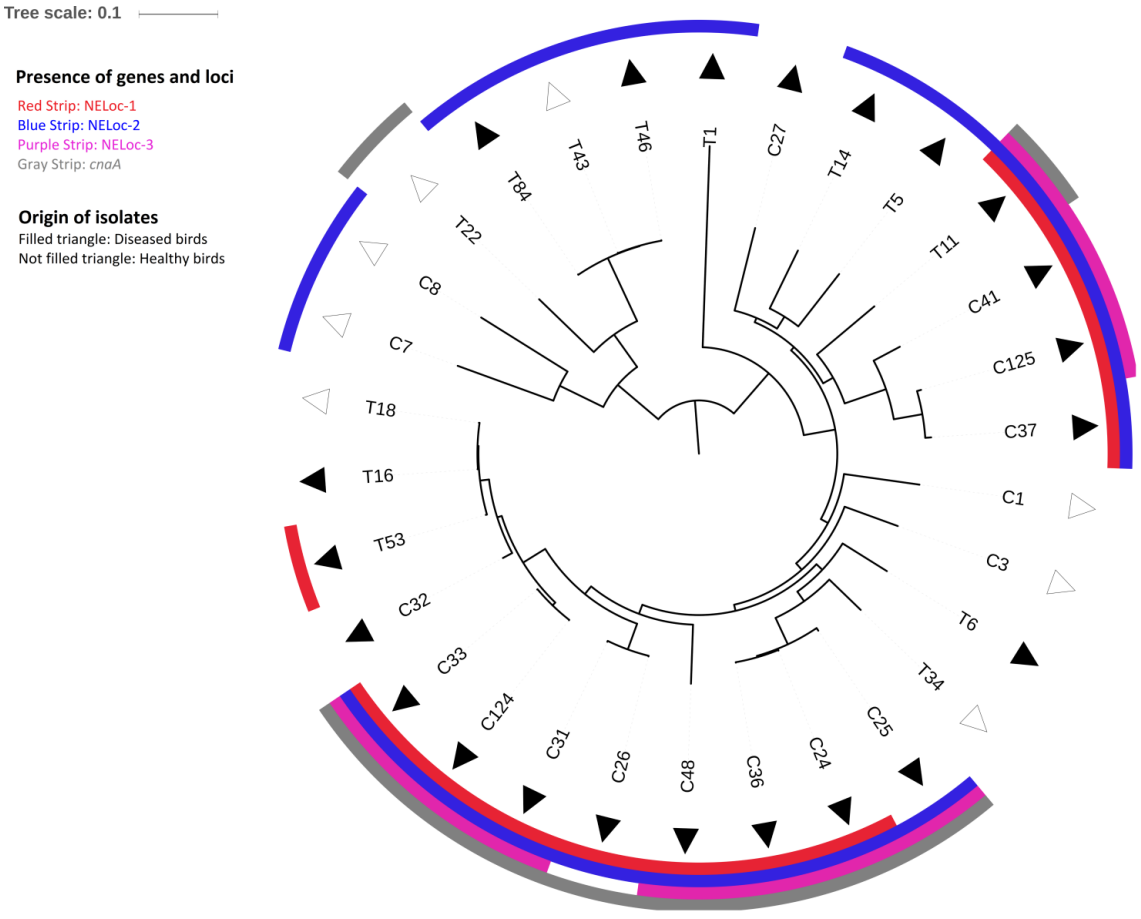
Isolate	State	Type	*cnaD*	*cnaA*	*netB*	NELoc-1	NELoc-2	NELoc-3
C1	H	A	+			0	0	0
C3	H	A				12	18	20
C7	H	A	+			3	**100**	20
C8	H	A				9	**100**	20
C24	D	A		+	+	**97**	**100**	**100**
C25	D	A		+		21	**100**	**100**
C26	D	A		+	+	**100**	**100**	**100**
C27	D	A	+			0	0	0
C31	D	A		+	+	**100**	**100**	**100**
C32	D	A	+			18	0	20
C33	D	A	+	+	+	**88**	**100**	**80**
C36	D	A		+	+	**100**	**100**	**100**
C37	D	A			+	**100**	**100**	60
C41	D	A			+	**100**	**100**	**100**
C48	D	A	+	+	+	**94**	**100**	**100**
C124	D	A	+	+	+	**100**	**100**	**80**
C125	D	A			+	**100**	**100**	**100**
T1	D	A	+			0	**100**	0
T5	D	A	+			27	**100**	20
T6	D	A	+			27	0	40
T11	D	A	+	+	+	**100**	**100**	**100**
T14	D	A	+			27	**100**	40
T16	D	A	+			9	0	20
T46	D	A	+			12	**100**	40
T53	D	A	+		+	**100**	0	40
T84	D	A	+			12	**100**	40
T18	H	A	+			21	0	60
T22	H	A		+		9	0	40
T34	H	A				9	0	20
T43	H	A	+			12	**100**	20

The table shows the presence and absence of two virulence genes (*cnaA* and *netB*) and a potential virulence gene, *cnaD*. The "+" mark indicates gene presence in isolates from healthy (H) or diseased (D) chickens (C) and turkeys (T). The prevalence (in %) of genes associated with three necrotic enteritis loci, NELoc-1, -2 and -3 is also shown. High prevalence of genes (≥ 80%) is highlighted in bold.

In previous studies *netB* was found in low prevalence, or was found to be totally absent, among NE isolates from turkeys whereas *netB* have been shown to be highly associated with NE in chickens (Lyhs et al., 2013)(Giovanardi et al., 2016)(Keyburn et al., 2010)(Keyburn et al., 2008). The three pathogenicity loci NELoc-1, -2 and -3 were initially identified in NE isolates from chickens (Lepp et al., 2010) which is why these loci in the present study also were considerably associated with NE isolates from chickens. Only NELoc-2 was found in high prevalence among NE isolates from turkeys suggesting that this locus play a role in regard to NE in turkeys. The *cnaA* gene has previously been identified exclusively among NE isolates from chickens (Wade et al., 2015) whereas another study also found it among healthy chickens (Lepp et al., 2013). Using comparative genomics in this study, a putative collagen adhesion gene was discovered among all isolates from diseased turkeys and has been designated *cnaD.* It encodes a von Willebrand Factor type A domain and three Cna protein B-

type domains which are types of domains that previously have been shown to be involved in collagen binding (Wade et al., 2015)(Jost et al., 2006)(Machha et al., 2016). This strongly indicates that CnaD is involved in collagen binding in the same way as CnaA. However, the function of CnaD should be further investigated and experimental studies as carried out previously by Wade et al., 2015 (Wade et al., 2015), could confirm more precisely which type of collagen CnaD has high affinity for. This could reveal if CnaD plays an essential role in the NE pathogenesis in turkeys. A phylogenetic tree based on SNPs showed no specific clustering pattern among the chicken and turkey isolates (Figure 4.2).

Figure 4.2 Phylogenetic tree of 30 *C perfringens* isolates and the distribution of virulence content



The maximum likelihood approximated tree shows that isolates from healthy and diseased chickens (C) and the turkeys (T) did not cluster into representative groups. NELoc-1,-3 and *cnaA* were significantly associated with NE isolates from chickens whereas NELoc-2 was associated with NE isolates from both chickens and turkeys. The scale bar represents substitutions per site.

Lastly, it should be mentioned that NE is a multifactorial disease and factors such as feed composition, mycotoxins, temperature and hygiene stress are involved in triggering of NE.

Therefore, it is not only the virulence gene content of pathogenic *C. perfringens* strains that is of importance when investigating NE outbreaks on poultry farms (Moore, 2016)(Lovland and Kaldhusdal, 2001).

Paper IV and V

In these two studies, isolates from Danish BTM and dairy cows with CM were whole-genome sequenced and further analyzed. To our knowledge, it is the first time that *S. aureus* isolates from these sources have been sequenced and made publicly available. All isolates were sequenced to an average coverage of > 50 fold except a single isolate that had 47 fold. In paper V statistical analyses showed no differences in presence of *spa*-types and STs between BTM and CM isolates except for ST97 and ST1 that were significantly associated with CM isolates (Table 4.3).

Table 4.3 Prevalence of STs and *spa*-types found among 157 *S. aureus* isolates

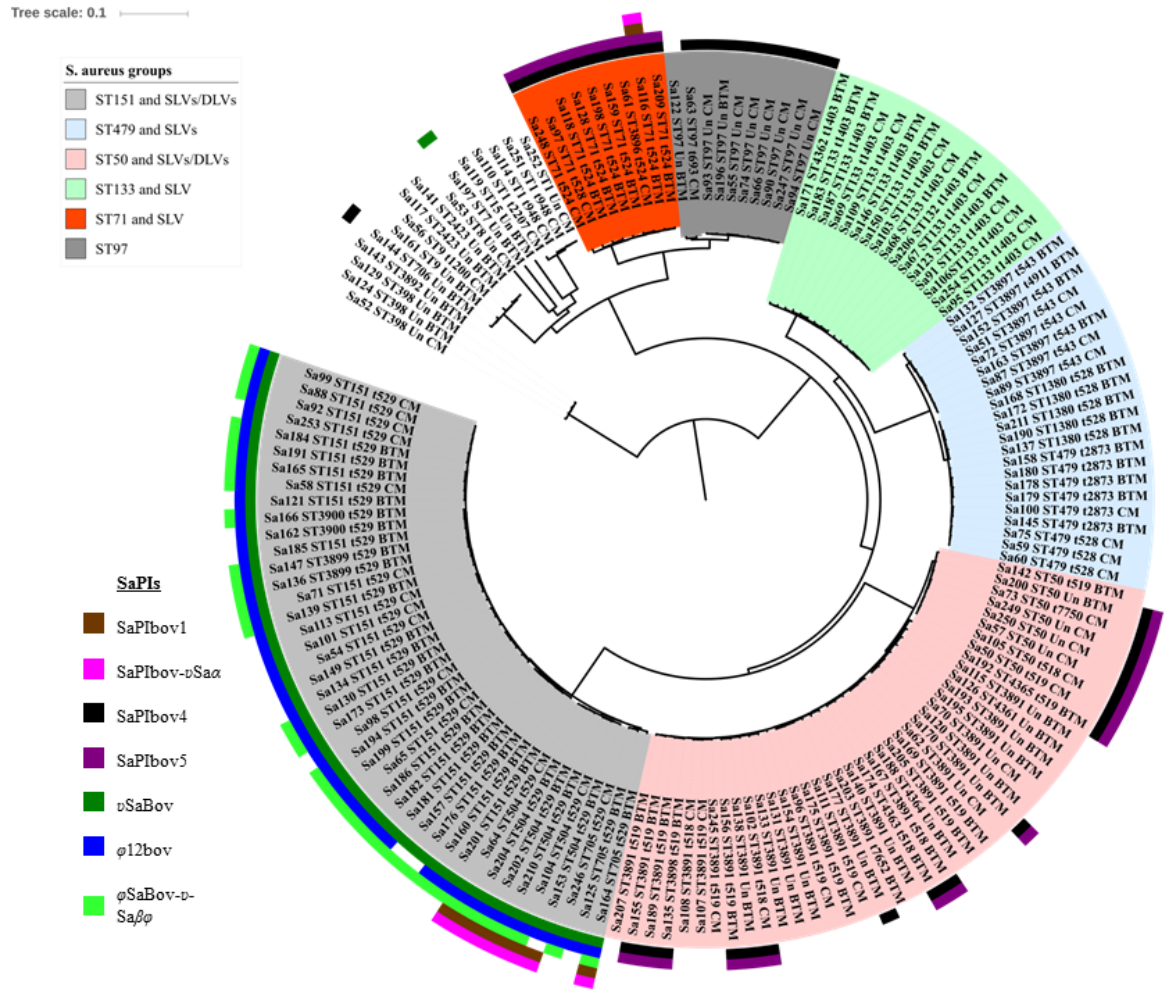
ST (%)	CM (n=63)	BTM (n=94)	*P*-value	*spa*-type (%)	CM (n=63)	BTM (n=94)	*P*-value
151 (19)	11	19	0.6672	**t529** (27)	14	29	0.2347
3891* (17)	8	19	0. 2213	**t519** (10)	6	10	0.8210
133 (9)	9	5	0.0838	**t1403** (10)	9	7	0.1650
97 (6)	8	2	*0.0151*	**t528** (6)	4	5	1.0000
479 (6)	4	5	1.0000	**t524** (5)	2	6	0.4768
50 (5)	6	2	0.0608	**t543** (5)	4	3	0.4395
71 (5)	2	6	0.4768	t2873 (4)	1	5	0.4027
3897* (5)	4	4	0.7146	t518 (3)	2	3	1.0000
504 (4)	2	4	1.0000	t693 (1)	1	0	0.4013
1 (3)	4	0	*0.0245*	t948 (1)	2	0	0.1595
1380 (3)	0	5	0.0831	t1200 (1)	1	0	0.4013
398 (2)	1	2	1.0000	t2207 (1)	1	0	0.4013
705 (2)	1	2	1.0000	t4911 (1)	0	1	1.0000
7 (1)	0	1	1.0000	t7652 (1)	0	1	1.0000
8 (1)	1	0	0.4013	t7750 (1)	1	0	0.4013
9 (1)	1	1	1.0000	Unk (25)	15	24	0.8066
15 (1)	0	1	1.0000				
132 (1)	0	1	1.0000				
706 (1)	0	1	1.0000				
2423 (1)	0	2	0.5164				
3892* (1)	0	1	1.0000				
3896* (1)	1	0	0.4013				
3898* (1)	0	2	0.5164				
3899* (1)	0	2	0.5164				
3900* (1)	0	2	0.5164				
4361* (1)	0	1	1.0000				
4362* (1)	0	1	1.0000				
4363* (1)	0	1	1.0000				
4364* (1)	0	1	1.0000				
4365* (1)	0	1	1.0000				

The prevalence of *spa*-types and STs among 63 clinical mastitis (CM) isolates and 94 isolates from bulk tank milk (BTM) is shown. The most prevalent *spa*-types and STs are highlighted in bold whereas new STs are marked with an asterisk (*). Differences in distributions of STs and *spa*-types between CM and BTM isolates were analyzed using statistical tests and significant *P*-values are shown in italics. Unk: Unknown.

However, only four ST1 isolates were found and therefore, it is difficult to conclude further on significance of this finding. The most prevalent STs were; ST151, 133, 97, 479, 50 and 71. Some of these STs (ST97, 133, 151 and 479) have previously been associated with mastitis (Holmes and Zadoks, 2011)(Boss et al., 2016), whereas ST50 and ST71 have been found among healthy cows and in BTM (Smith et al., 2005)(Hata et al., 2010). Moreover, 12 new STs were observed and registered in the PubMLST database (Jolley et al., 2004). The two most prevalent of the new STs were ST3891 and ST3897 which were single locus variants of ST50 and ST479, respectively. BTM isolates may originate from the extra-mammary sites such as teat skin, teat canal or milking staff but also from subclinical infected quarters (Haveri et al., 2008). This could be the reason why a more diverse composition of STs was observed among BTM samples (22 different STs) compared to CM samples (15 different STs). The most prevalent *spa*-type, t529 was observed in 27% of the isolates, followed by t1403 and t519 that were both found in 10% of the isolates. These three *spa*-types have all been associated with bovine mastitis but also healthy cows (Boss et al., 2016)(Ikawaty et al., 2009)(Sakwinska et al., 2011) (Johler et al., 2011)(Hasman et al., 2010). Twenty-five % of the isolates were identified as unknown *spa*-types by spaTyper (Bartels et al., 2014). Analyses revealed that the *spa* genes were located on > 1 contig and therefore were not identified by spaTyper which only register *spa*-genes entirely conserved on a single contig (Bartels et al., 2014). Assembly and sequencing error could also explain why the order of the *spa* repeats was not determined correctly. However, it was beyond the scope of this study to identify and register new *spa*-types according to the official guidelines provided at http://spaserver.ridom.de/.

Moreover, a phylogenetic tree showed that BTM and CM isolates clustered together into groups of closely related STs (Figure 4.3).

Figure 4.3 Phylogenetic tree of 157 *S aureus* isolates including the distribution of SaPIs



The maximum-likelihood approximated tree that includes 63 clinical mastitis (CM) isolates and 94 isolates from bulk tank milk (BTM), shows that isolates from BTM and CM cluster together into groups of closely related STs. Three specific SaPIs (*φ*12bov, υSaBov and φSaBov-υ-Saβφ) were almost only found in a cluster that primarily consisted of ST151 isolates (marked in light gray). In contrast, SaPIbov4 and SaPIbov5 were found among various STs that were not all closely related.

A large cluster of isolates that primarily belonged to ST151 was observed. The vast majority of the isolates in this group carried three SaPIs (*φ*12bov, *υ*SaBov and *φ*SaBov-*υ*-Sa*βφ*) found in strain RF122 (Herron-Olson et al., 2007). These three SaPIs were only found in this cluster except for a single ST7 isolate that carried υSaBov. Strain RF122 has previously been described as a commonly observed mastitis causing clone type (Fritzgerald et al., 1997). The three SaPIs originating from RF122 carries a variety of virulence genes and therefore it has been suggested that these SaPIs are important regarding the CM pathogenesis and successful adaption to dairy cows (Herron-Olson et al., 2007). Notably, both statistical and phylogenetic analysis showed that the BTM and CM isolates in general were of identical genetic

background. These findings suggest that dairy cows are natural carriers of *S. aureus* subtypes that cause CM. These subtypes can cause CM if for example the cows appear immunocompromised and poor hygiene occurs. Concordantly, it has previously been reported that STs and *spa*-types that were often associated with bovine mastitis also can be present in healthy cows and BTM (Boss et al., 2016)(Conceicão, 2017)(Jørgensen et al., 2005).

When looking at the presence of antibiotic resistance and virulence genes there were no statistical differences between BTM and CM isolates in regard to resistance genes. However, five enterotoxin genes (*sei*, *sem*, *sen*, *seo* and *seu*) were significantly associated with BTM isolates whereas a serine protease gene (*splE*) and an enterotoxin gene (*seh*) were significantly associated with CM isolates (Table 4.4).

Table 4.4 Prevalence of virulence and resistance genes among 157 *S. aureus* isolates

Virulence genes (%)	CM (n=63)	BTM (n=94)	*P*-value	Resistance genes (%)	CM (n=63)	BTM (n=94)	*P*-value
aur (100)	63	94	1.0000	*norA* (99)	62	94	0.4013
hla (100)	63	94	1.0000	*blaZ* (17)	13	14	0.3501
hlb (99)	63	93	1.0000	*tetM* (3)	1	3	0.6495
hlgB (99)	63	93	1.0000	*dfrG* (2)	0	3	0.2746
hlgC (96)	61	89	0.7027	*ermB* (1)	1	0	0.4013
fib (96)	59	92	0.2196	*lnuA* (1)	0	1	1.0000
nuc (95)	61	88	0.4768	*lnuB* (1)	1	1	1.0000
icaD (94)	61	87	0.3162	*mecA* (1)	1	0	0.4013
hlgA (94)	58	89	0.5236	*tetK* (1)	1	0	0.4013
splA (92)	60	85	0.3638	*vgaA* (1)	0	1	1.0000
splB (92)	60	85	0.3638				
lukD (89)	55	85	0.7296				
lukE (81)	49	78	0.4165				
seu (69)	36	72	*0.0099*				
sem (68)	34	72	*0.0030*				
sen (68)	36	70	*0.0231*				
seo (66)	33	71	*0.0026*				
sei (66)	34	69	*0.0120*				
seg (45)	23	47	0.0955				
splE (16)	15	10	*0.0270*				
sec (5)	2	5	0.7027				
sel (5)	2	5	0.7027				
tst (5)	2	5	0.7027				
scn (3)	3	1	0.3029				
seh (3)	4	0	*0.0245*				
seq (2)	3	0	0.0628				
sak (2)	3	0	0.0628				
sek (2)	3	0	0.0628				
sea/sep (2)	3	0	0.0628				

The prevalence of 29 virulence genes and 10 resistance genes among 63 clinical mastitis (CM) isolates and 94 isolates from bulk tank milk (BTM) is shown. The virulence genes could be divided into three groups: **Group 1** (genes found in ≥ 81%), **Group 2** (genes found in 45-69%) and **Group 3** (genes found in 2-16%). Differences in distributions of virulence and resistance genes between CM and BTM isolates were investigated using statistical tests and significant *P*-values are shown in italics.

The *seh* gene was found only in four isolates and it is therefore difficult to conclude anything from this result. Previously, enterotoxin genes have been found among bovine mastitis isolates (Fueyo et al., 2005)(Xu et al., 2015)(Fournier et al., 2008)(Kot et al., 2016) but the role of enterotoxins in the pathogenesis is not fully elucidated. Studies suggest that they are not essential in the mastitis pathogenesis (Larsen et al., 2002)(Larsen et al., 2000) and a previous investigation of 106 Danish isolates associated with subclinical mastitis showed that none of the isolates carried any enterotoxin genes (Aarestrup et al., 1995b). Since enterotoxins are heat-stable they may be present in various dairy products such as milk even after heat treatment (Jørgensen et al., 2005)(Hennekinne et al., 2012). Therefore, enterotoxins have been reported to be associated with staphylococcal food poisoning caused by cow milk or other dairy products (Asao et al., 2003)(Jørgensen et al., 2005)(Hennekinne et al., 2012). In contrast to the enterotoxin genes, a single serine protease gene (*spl*E) was significantly associated with CM isolates. Previously, serine proteases have been described as an important virulence factor and protease genes have been observed in high prevalence among mastitis isolates (Kot et al., 2016).

The prevalence of resistance genes was in general low even though *norA* was found in almost all isolates and *blaZ* was found in 17% of the isolates. Concordantly, a previous Danish study found that 17% of 105 isolates produced beta-lactamases (Aarestrup et al., 1995a). *norA* mediates resistance to quinolones and other antiseptic compounds (Santos Costa et al., 2015)(Kaatz and Seo, 1995). Danish dairy cows with CM are not treated with quinolones and thus the presence of *norA* must be associated with other factors (www.foedevarestyrelsen.dk/Leksikon/Sider/VetStat.aspx). It could be speculated if *norA* is involved in mediating resistance against antiseptic compounds used in the Danish dairy industry to increase the hygiene. Furthermore, the first ST398 MRSA isolate (Sa52) from a Danish dairy cow with CM was discovered. Strain Sa52 carried many different resistance genes compared to the remaining 62 CM isolates that only carried *blaZ* and *norA*. In paper IV, the assemblies of strain Sa52 was annotated and uploaded to NCBI. The draft genome consisted of 145 contigs and had an assembly size of 2.8 Mb and an average coverage of 103 fold. The annotation resulted in a total of 2878 genes, where of 2735 were predicted to be protein-coding sequences. The *mecA*-positive strain Sa52 carried the staphylococcal cassette chromosome *mec* element V(5C2&5)c (Vandendriessche et al., 2014) and other virulence genes (*aur*, *cap*, *fib*, *hla*, *hlb*, *hlg*, *icaD* and *nuc*) that have been associated with bovine

mastitis (Xu et al., 2015)(Bardiau et al., 2016). Interestingly, strain Sa52 carried no functional restriction modification system which means that foreign DNA can easily be taken up (Murray, 2000). Previous studies suggest that the ST398 lineage has the ability to jump between pigs and humans and thereby develop into a serious human pathogen (Larsen et al., 2015)(Price et al., 2012). Successful host-shifts could require uptake of foreign DNA and thus a weak restriction enzyme barrier will potentially favors these host-shifts. However, the transmission of *S. aureus* strains between humans and dairy cows seems to occur less frequently (Boss et al., 2016)(Sakwinska et al., 2011). In the livestock industry MRSA strains belonging to ST398 have primarily been associated with pigs but strain Sa52 was sampled from a farm where no pig farming had taken place (Larsen et al., 2015)(Price et al., 2012). Therefore, it is unlikely that it was directly related to pig farming but could have been transmitted by a visitor or a farm worker who carried the clone.

5. Conclusions and future perspectives

Paper I and II

Investigation of APEC isolates from broiler chickens and parents with colibacillosis, collected in different Nordic countries, showed a group of 47 closely related ST117 O78:H4 isolates that also shared the same virulence profile. These results suggest that the ST117 O78:H4 isolates have been transmitted vertically through the broiler breeding pyramid and constituted the main reason for the increased mortality rates observed on Nordic poultry farms between 2014 and 2016. In Denmark, a new autogenous vaccine program introduced in 2015 included strain E44 that belonged to the group of ST117 O78:H4 isolates. After this introduction, both the number of colibacillosis cases and the mortality rate have decreased considerably (Karl Pedersen, personal communication). In contrast, autogenous vaccines have not been used in Finland and an increased mortality rate is still observed (Tarja Pohjanvirta, personal communication). Therefore, it seems like autogenous vaccines can be used efficiently to avoid colibacillosis outbreaks on poultry farms if of course, it protects against the right pathogens.

In paper I, assemblies from the two APEC strains (E44 and E51) that have been used in a Danish autogenous *E. coli* vaccine program, were annotated and added to the NCBI database. Thus, these two draft genomes are now publicly available and can be used for comparative studies in the future.

Paper III

In this study, *C. perfringens* isolates from healthy and NE afflicted chickens and turkeys were whole-genome sequenced and downstream analyses carried out. It is probably the first time that *C. perfringens* isolates from turkeys with NE have been whole-genome sequenced and made publicly available. Thus, these sequences can potentially be very valuable in future studies that address NE in turkeys. It was shown that specific pathogenicity loci and virulence genes (NELoc-1, -3, *netB* and *cnaA*) were significantly associated with NE isolates from chickens compared to those from turkeys. Only NELoc-2 was associated with NE isolates from both turkeys and chickens. This suggests that the NE pathogenesis in chickens is different from that of turkeys. NE is triggered by many different disease factors such as feed composition, mycotoxins, temperature and hygiene stress. Thus, it is important for farmers to be aware of these disease factors in order to avoid NE outbreaks on the farms. Microbiome studies could reveal which potential impact these factors have on the composition of the gut

microbiota in both healthy and NE infected turkeys and chickens. Additionally, metagenomics studies could potentially elucidate the single strain dominance phenomenon in diseased turkeys. Lastly, a putative collagen adhesion gene designated *cnaD* was discovered in all diseased turkeys. Potentially, *cnaD* could be of importance regarding the NE pathogenesis in turkeys but needs further investigation.

Paper IV and V

In these studies, Danish *S. aureus* isolates from BTM and CM were whole-genome sequenced and analyzed in regard to genomic content and genetic relationship. Presumably, it is the first time that such types of isolates from Denmark have been whole-genomes sequenced and made publicly available.

In paper V statistical and phylogenetic analyses showed that isolates from both BTM and CM generally were of similar genetic background. This suggests that dairy cows can be natural carriers of *S. aureus* subtypes that can cause clinical mastitis if the right conditions are present. A phylogenetic tree showed a large cluster that primarily consisted of ST151 isolates. Isolates from this cluster carried three SaPIs that were almost only found in this group and probably are involved in the mastitis pathogenesis and host adaption. The prevalence of antibiotic resistance genes was in general low but the first ST398 MRSA isolate from a Danish dairy cow with CM was observed. The ST398 lineage has the ability to jump between humans and livestock and thus it is important to further monitor cattle herds to avoid future problems regarding MRSA strains with zoonotic potential. In paper IV, assemblies of this MRSA isolate, strain Sa52 was annotated and uploaded to NCBI. Thus, it has been made publicly available and can therefore easily be used for analyses in the future.

Final remarks

In this PhD project, whole-genome sequencing was applied to analyze various types of important veterinary pathogens. Phylogenetic analyses based SNPs were carried out and unique data obtained. Specifically, the phylogenetic analysis presented in paper II Fig 4.1 was able to differentiate between closely related APEC strains that belonged to the same ST and serotype (ST117 O78:H4). This type of analysis could not have been generated using traditional molecular techniques such as PFGE since the discriminatory power is simply too low. Thus, SNP-based phylogenetic analysis is a valuable tool to trace outbreaks and to

investigate the epidemiology of closely related subtypes. Whole-genome sequencing is also a very useful tool in regard to population dynamics. For example, phylogenic studies suggest that methicillin sensitive ST398 clones from humans were introduced into the livestock environment. Here, the ST398 clones became methicillin resistant on multiple occasions, spread among farms and are now, due to repeated spill over, causing infections in not only farmers but also in the general public (Price et al., 2012)(Larsen et al., 2015).

Furthermore, a potential virulence gene, *cnaD* was identified in paper III and it would probably have been more laborious, time consuming and expensive to identify this gene via experimental laboratory work. Thus, whole-genome sequencing is an excellent tool for initial identification and investigation of potential virulence factors in pathogenic bacteria and for selection of potential veterinary autogenous vaccine candidates. However, novel *in silico* data sometimes needs to be combined with experimental results such as in the case of confirming the precise function of the potential new virulence factor, CnaD. In general, whole-genome sequencing is a very useful technique within many areas of veterinary microbiology even though it is not an error-free technique. Additionally, DNA sequencing techniques are developing remarkably fast and will probably be much faster and cost effective in the future. Thus, it is assumed that whole-genome sequencing will be widely implemented in many fields of microbiology in the future since it provides various unique advantages compared to traditional molecular techniques.

References

Adiri, R.S., Gophna, U., Ron, E.Z., 2003. Multilocus sequence typing (MLST) of *Escherichia coli* O78 strains. FEMS Microbiol. Lett. 222, 199–203. doi:10.1016/S0378-1097(03)00295-7

Agersø, Y., Hasman, H., Cavaco, L.M., Pedersen, K., Aarestrup, F.M., 2012. Study of methicillin resistant *Staphylococcus aureus* (MRSA) in Danish pigs at slaughter and in imported retail meat reveals a novel MRSA type in slaughter pigs. Vet. Microbiol. 157, 246–250. doi:10.1016/j.vetmic.2011.12.023

Agersø, Y., Jensen, J.D., Hasman, H., Pedersen, K., 2014. Spread of extended spectrum cephalosporinase-producing *Escherichia coli* clones and plasmids from parent animals to broilers and to broiler meat in a production without use of cephalosporins. Foodborne Pathog. Dis. 11, 740–746. doi:10.1089/fpd.2014.1742

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST:a new generation of protein database search programs. Nucleic Acids Res 25, 3389–3402. doi:10.1093/nar/25.17.3389

Angiuoli, S. V, Gussman, A., Klimke, W., Cochrane, G., Field, D., Garrity, G.M., Kodira, C.D., Kyrpides, N., Madupu, R., Markowitz, V., Tatusova, T., Thomson, N., White, O., 2008. Toward an Online Repository of Standard Operating Procedures (SOPs) for (Meta)genomic Annotation. Omi. A J. Integr. Biol. 12, 137–141. doi:10.1089/omi.2008.0017

Asao, T., Kumeda, Y., Kawai, T., Shibata, T., Oda, H., Haruki, K., Nakazawa, H., Kozaki, S., 2003. An extensive outbreak of staphylococcal food poisoning due to low-fat milk in Japan: estimation of enterotoxin A in the incriminated milk and powdered skim milk. Epidemiol. Infect. 130, 33–40. doi:10.1017/S0950268802007951

Barbara, A.J., Trinh, H.T., Glock, R.D., Glenn Songer, J., 2008. Necrotic enteritis-producing strains of *Clostridium perfringens* displace non-necrotic enteritis strains from the gut of chicks. Vet. Microbiol. 126, 377–382. doi:10.1016/j.vetmic.2007.07.019

Bardiau, M., Caplin, J., Detilleux, J., Graber, H., Moroni, P., Taminiau, B., Mainil, J.G., 2016. Existence of two groups of Staphylococcus aureus strains isolated from bovine mastitis based on biofilm formation, intracellular survival, capsular profile and *agr*-typing. Vet. Microbiol. 185, 1–6. doi:10.1016/j.vetmic.2016.01.003

Bartels, M.D., Petersen, A., Worning, P., Nielsen, J.B., Larner-Svensson, H., Johansen, H.K.,

Andersen, L.P., Jarløv, J.O., Boye, K., Larsen, A.R., Westh, H., 2014. Comparing whole-genome sequencing with sanger sequencing for spa typing of methicillin-resistant *Staphylococcus aureus*. J. Clin. Microbiol. 52, 4305–4308. doi:10.1128/JCM.01979-14

Bisgaard M, Pedersen JR, Christensen H. Danish Commercial Poultry (Dansk Erhvervsfjerkræ), Oct 2015. 10: 14-19, n.d.

Boss, R., Cosandey, A., Luini, M., Artursson, K., Bardiau, M., Al, E., 2016. Bovine *Staphylococcus aureus* : Subtyping , evolution , and zoonotic transfer. J. Dairy Sci. 515–528.

Chen, L., Zheng, D., Liu, B., Yang, J., Jin, Q., 2016. VFDB 2016: hierarchical and refined dataset for big data analysis-10 years on. Nucleic Acids Res. 44, D694-7. doi:10.1093/nar/gkv1239

Cheung, J.K., Keyburn, A.L., Carter, G.P., Lanckriet, A.L., Van Immerseel, F., Moore, R.J., Rood, J.I., 2010. The VirSR two-component signal transduction system regulates NetB toxin production in *Clostridium perfringens*. Infect. Immun. 78, 3064–3072. doi:10.1128/IAI.00123-10

Cheung, J.K., Awad, M.M., McGowan, S., Rood, J.I., 2009. Functional analysis of the VirSR phosphorelay from Clostridium perfringens. PLoS One 4, e5849. doi:10.1371/journal.pone.0005849

Conceicão, T., 2017. Healthy Bovines as Reservoirs of Major Pathogenic Lineages of *Staphylococcus aureus* in Portugal. Microb. drug Resist. 0, 1–7. doi:10.1089/mdr.2017.0074

Croucher, N.J., Page, A.J., Connor, T.R., Delaney, A.J., Keane, J.A., Bentley, S.D., Parkhill, J., Harris, S.R., 2015. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res 43. doi:10.1093/nar/gku1196

Deurenberg, R.H., Nieuwenhuis, R.F., Driessen, C., London, N., Stassen, F.R., Tiel, F.H. Van, Stobberingh, E.E., Vink, C., 2005. The prevalence of the *Staphylococcus aureus tst* gene among community- and hospital-acquired strains and isolates from Wegener's Granulomatosis patients. FEMS Microbiol. Lett. 245, 185–189. doi:10.1016/j.femsle.2005.03.002

Dho-moulin, M., Fairbrother, J.M., 1999. Avian pathogenic *Escherichia coli* (APEC). Vet. Microbiol. 30, 299–316.

Didelot, X., Bowden, R., Wilson, D.J., Peto, T.E.A., Crook, D.W., 2012. Transforming clinical microbiology with bacterial genome sequencing. Nat. Rev. Genet. 13, 601–612. doi:10.1038/nrg3226

Dissanayake, D.R.A., Octavia, S., Lan, R., 2014. Population structure and virulence content of avian pathogenic *Escherichia coli* isolated from outbreaks in Sri Lanka. Vet. Microbiol. 168, 403–412. doi:10.1016/j.vetmic.2013.11.028

Dozois, C.M., Fairbrother, J.M., Harel, J., Bossé, M., 1992. *pap*-and *pil*-related DNA sequences and other virulence determinants associated with *Escherichia coli* isolated from septicemic chickens and turkeys. Infect. Immun. 60, 2648–56.

Earl, D., Bradnam, K., St. John, J., Darling, A., Lin, D., Fass, J., Yu, H.O.K., Buffalo, V., Zerbino, D.R., Diekhans, M., Nguyen, N., Ariyaratne, P.N., Sung, W.K., Ning, Z., Haimel, M., Simpson, J.T., Fonseca, N.A., Birol, I., Docking, T.R., Ho, I.Y., Rokhsar, D.S., Chikhi, R., Lavenier, D., Chapuis, G., Naquin, D., Maillet, N., Schatz, M.C., Kelley, D.R., Phillippy, A.M., Koren, S., Yang, S.P., Wu, W., Chou, W.C., Srivastava, A., Shaw, T.I., Ruby, J.G., Skewes-Cox, P., Betegon, M., Dimon, M.T., Solovyev, V., Seledtsov, I., Kosarev, P., Vorobyev, D., Ramirez-Gonzalez, R., Leggett, R., MacLean, D., Xia, F., Luo, R., Li, Z., Xie, Y., Liu, B., Gnerre, S., MacCallum, I., Przybylski, D., Ribeiro, F.J., Sharpe, T., Hall, G., Kersey, P.J., Durbin, R., Jackman, S.D., Chapman, J.A., Huang, X., DeRisi, J.L., Caccamo, M., Li, Y., Jaffe, D.B., Green, R.E., Haussler, D., Korf, I., Paten, B., 2011. Assemblathon 1: A competitive assessment of de novo short read assembly methods. Genome Res. 21, 2224–2241. doi:10.1101/gr.126599.111

Ewers, C., Li, G., Wilking, H., Kießling, S., Alt, K., Antáo, E.M., Laturnus, C., Diehl, I., Glodde, S., Homeier, T., Böhnke, U., Steinrück, H., Philipp, H.C., Wieler, L.H., 2007. Avian pathogenic, uropathogenic, and newborn meningitis-causing *Escherichia coli*: How closely related are they? Int. J. Med. Microbiol. 297, 163–176. doi:10.1016/j.ijmm.2007.01.003

Foley, S.L., Zhao, S., Walker, R.D., 2007. Comparison of Molecular Typing Methods for the Differentiation of Salmonella Foodborne Pathogens. Foodborne Pathog. Dis. 4.

Fournier, C., Kuhnert, P., Frey, J., Miserez, R., Kirchhofer, M., Kaufmann, T., Steiner, A., Graber, H.U., 2008. Bovine *Staphylococcus aureus*: Association of virulence genes, genotypes and clinical outcome. Res. Vet. Sci. 85, 439–448. doi:10.1016/j.rvsc.2008.01.010

Fritzgerald, J.R., Meaney, W.J., Hartigan, P.J., Smyth, C.J., Kapur, V., 1997. Fine-structure molecular epidemiological analysis of *Staphylococcus aureus* recovered from cows. Epidemiol. Infect. 261–269.

Fueyo, J.M., Mendoza, M.C., Rodicio, M.R., Muñiz, J., Alvarez, M.A., Martín, M.C., 2005. Cytotoxin and pyrogenic toxin superantigen gene profiles of *Staphylococcus aureus* associated with subclinical mastitis in dairy cows and relationships with macrorestriction genomic profiles. J. Clin. Microbiol. 43, 1278–1284. doi:10.1128/JCM.43.3.1278-1284.2005

Giovanardi, D., Drigo, I., De Vidi, B., Agnoletti, F., Viel, L., Capello, K., Berto, G., Bano, L., 2016. Molecular characterization of Clostridium perfringens strains isolated from diseased turkeys in Italy. Avian Pathol. 9457, 1–17. doi:10.1080/03079457.2016.1160364

Goodyear, C.S., Silverman, G.J., 2003. Death by a B Cell Superantigen : In Vivo V H -targeted Apoptotic Supraclonal B Cell Deletion by a Staphylococcal Toxin. J. Exp. Med. 197. doi:10.1084/jem.20020552

Halasa, T., Huijps, K., Østerås, O., Hogeveen, H., 2007. Economic effects of bovine mastitis and mastitis management: A review. Vet. Q. 29, 18–31. doi:10.1080/01652176.2007.9695224

Harmsen, D., Claus, H., Witte, W., Claus, H., Turnwald, D., Vogel, U., 2003. Typing of Methicillin-Resistant Staphylococcus aureus in a University Hospital Setting by Using Novel Software for spa Repeat Determination and Database Management. J. Clin. Microbiol. 41, 5442–5448. doi:10.1128/JCM.41.12.5442

Hasman, H., Moodley, A., Guardabassi, L., Stegger, M., Skov, R.L., Aarestrup, F.M., 2010. *spa* type distribution in *Staphylococcus aureus* originating from pigs, cattle and poultry. Vet. Microbiol. 141, 326–331. doi:10.1016/j.vetmic.2009.09.025

Hasman, H., Saputra, D., Sicheritz-Ponten, T., Lund, O., Svendsen, C.A., Frimodt-Moller, N., Aarestrup, F.M., 2014. Rapid whole-genome sequencing for detection and characterization of microorganisms directly from clinical samples. J. Clin. Microbiol. 52, 139–146. doi:10.1128/JCM.02452-13

Hata, E., Katsuda, K., Kobayashi, H., Uchida, I., Tanaka, K., Eguchi, M., 2010. Genetic variation among *Staphylococcus aureus* strains from bovine milk and their relevance to methicillin-resistant isolates from humans. J. Clin. Microbiol. 48, 2130–2139.

doi:10.1128/JCM.01940-09

Haveri, M., Hovinen, M., Roslo, A., Pyo, S., 2008. Molecular types and genetic profiles of *Staphylococcus aureus* strains isolated from bovine intramammary infections and extramammary sites. J. Clin. Microbiol. 46, 3728–3735. doi:10.1128/JCM.00769-08

Hennekinne, J., Buyser, M. De, Dragacci, S., 2012. *Staphylococcus aureus* and its food poisoning toxins: characterization and outbreak investigation. Fems Microb. Rev 36, 815–836. doi:10.1111/j.1574-6976.2011.00311.x

Herron-olson, L., Fitzgerald, J.R., Musser, J.M., Kapur, V., 2007. Molecular Correlates of Host Specialization in *Staphylococcus aureus*. doi:10.1371/journal.pone.0001120

Holmes, M.A., Zadoks, R.N., 2011. Methicillin resistant *S. aureus* in human and bovine mastitis. J. Mammary Gland Biol. Neoplasia 16, 373–382. doi:10.1007/s10911-011-9237-x

Ikawaty, R., Brouwer, E.C., Jansen, M.D., Duijkeren, E. Van, Mevius, D., Verhoef, J., Fluit, A.C., 2009. Characterization of Dutch *Staphylococcus aureus* from bovine mastitis using a Multiple Locus Variable Number Tandem Repeat Analysis. Vet. Microbiol. 136, 277–284. doi:10.1016/j.vetmic.2008.10.034

Joensen, K.G., Scheutz, F., Lund, O., Hasman, H., Kaas, R.S., Nielsen, E.M., Aarestrup, F.M., 2014. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. J. Clin. Microbiol. 52, 1501–1510. doi:10.1128/JCM.03617-13

Joensen, K.G., Tetzschner, A.M.M., Iguchi, A., Aarestrup, F.M., Scheutz, F., 2015. Rapid and easy *in silico* serotyping of *Escherichia coli* isolates by Use of whole-genome sequencing data. J. Clin. Microbiol. 53, 2410–2426. doi:10.1128/JCM.00008-15

Johler, S., Layer, F., Stephan, R., 2011. Comparison of Virulence and Antibiotic Resistance Genes of Food Poisoning Outbreak Isolates of *Staphylococcus aureus* with Isolates Obtained from Bovine Mastitis Milk and Pig Carcasses. J. Food Prot. 74, 1852–1859. doi:10.4315/0362-028X.JFP-11-192

Jolley, K.A., Chan, M.-S., Maiden, M.C.J., 2004. mlstdbNet - distributed multi-locus sequence typing (MLST) databases. BMC Bioinformatics 5, 86. doi:10.1186/1471-2105-5-86

Jordan, F.T.W., Williams, N.J., Wattret, a, Jones, T., 2005. Observations on salpingitis, peritonitis and salpingoperitonitis in a layer breeder flock. Vet. Rec. 157, 573–577.

doi:10.1136/vr.157.19.573

Jost, B.H., Billington, S.J., Trinh, H.T., Songer, J.G., 2006. Association of genes encoding beta2 toxin and a collagen binding protein in *Clostridium perfringens* isolates of porcine origin. Vet. Microbiol. 115, 173–182. doi:10.1016/j.vetmic.2006.01.012

Jørgensen, H.J., Mørk, T., Rørvik, L.M., 2005. The Occurrence of *Staphylococcus aureus* on a Farm with Small-Scale Production of Raw Milk Cheese. J. Dairy Sci. 88, 3810–3817. doi:10.3168/jds.S0022-0302(05)73066-6

Katholm, J., Bennedsgaard, T.W., Koskinen, M.T., Rattenborg, E., 2012. Quality of bulk tank milk samples from Danish dairy herds based on real-time polymerase chain reaction identification of mastitis pathogens. J. Dairy Sci. 95, 5702–8. doi:10.3168/jds.2011-5307

Keyburn, A.L., Boyce, J.D., Vaz, P., Bannam, T.L., Ford, M.E., Parker, D., Di Rubbo, A., Rood, J.I., Moore, R.J., 2008. NetB, a new toxin that is associated with avian necrotic enteritis caused by *Clostridium perfringens*. PLoS Pathog. 4. doi:10.1371/journal.ppat.0040026

Keyburn, A.L., Yan, X.X., Bannam, T.L., Van Immerseel, F., Rood, J.I., Moore, R.J., 2010. Association between avian necrotic enteritis and *Clostridium perfringens* strains expressing NetB toxin. Vet. Res. 41, 21. doi:10.1051/vetres/2009069

Kot, B., Szweda, P., Frankowska-Maciejewska, A., Piechota, M., Wolska, K., 2016. Virulence gene profiles in *Staphylococcus aureus* isolated from cows with subclinical mastitis in eastern Poland. J. Dairy Res. 83, 228–235. doi:10.1017/S002202991600008X

Kaas, R.S., Leekitcharoenphon, P., Aarestrup, F.M., Lund, O., 2014. Solving the problem of comparing whole bacterial genomes across different sequencing platforms. PLoS One 9, 1–8. doi:10.1371/journal.pone.0104984

Kaatz, G.W., Seo, S.M., 1995. Inducible NorA-Mediated Multidrug Resistance in *Staphylococcus aureus*. Antimicrob. Agents Chemother. 39, 2650–2655.

Landman, W.J., van Eck, J.H., 2015. The incidence and economic impact of the *Escherichia coli* peritonitis syndrome in Dutch poultry farming. Avian Pathol. 9457, 1–27. doi:10.1080/03079457.2015.1060584 [doi]

Larsen, H.D., Sloth, K.H., Elsberg, C., Enevoldsen, C., 2000. The dynamics of *Staphylococcus aureus* intramammary infection in nine Danish dairy herds. Vet. Microbiol. 71, 89–101.

Larsen, H.D., Aarestrup, F.M., Jensen, N.E., 2002. Geographical variation in the presence of

genes encoding superantigenic exotoxins and beta-hemolysin among *Staphylococcus aureus* isolated from bovine mastitis in Europe and USA. Vet. Microbiol. 85, 61–67.

Larsen, J., Petersen, A., Sørum, M., Stegger, M., Van Alphen, L., Valentiner-Branth, P., Knudsen, L.K., Larsen, L.S., Feingold, B., Price, L.B., Andersen, P.S., Larsen, A.R., Skov, R.L., 2015. Meticillin-resistant *staphylococcus aureus* CC398 is an increasing cause of disease in people with no livestock contact in Denmark, 1999 to 2011. Eurosurveillance 20. doi:10.2807/1560-7917.ES.2015.20.37.30021

Larsen, M.V., Joensen, K.G., Zankari, E., Ahrenfeldt, J., Lukjancenko, O., Kaas, R.S., Roer, L., Leekitcharoenphon, P., Saputra, D., Cosentino, S., Christen, M., Thomsen, F., Luis, J., Cisneros, B., Jurtz, V., Rasmussen, S., Petersen, T.N., Hasman, H., Sicheritz-ponten, T., Aarestrup, F.M., Lund, O., 2017. The CGE Tool Box. Appl. Genomics Foodborne Pathog. doi:10.1007/978-3-319-43751-4

Larsen, Huda, A., Eriksen, N.H.R., Jensen, N.E., 2000. Differences between Danish bovine and human *Staphylococcus aureus* isolates in possession of superantigens. Vet. Microbiol. 76, 153–162.

Larsen, M. V., Cosentino, S., Rasmussen, S., Friis, C., Hasman, H., Marvig, R.L., Jelsbak, L., Sicheritz-Ponten, T., Ussery, D.W., Aarestrup, F.M., Lund, O., 2012. Multilocus Sequence Typing of Total-Genome-Sequenced Bacteria. J. Clin. Microbiol. 50, 1355–1361. doi:10.1128/JCM.06094-11

Leekitcharoenphon, P., Kaas, R.S., Christen, M., Thomsen, F., Friis, C., Rasmussen, S., Aarestrup, F.M., 2012. snpTree - a web-server to identify and construct SNP trees from whole genome sequence data. BMC Genomics 13, 1–8.

Lepp, D., Gong, J., Songer, J.G., Boerlin, P., Parreira, V.R., Prescott, J.F., 2013. Identification of accessory genome regions in poultry *clostridium perfringens* isolates carrying the *netB* plasmid. J. Bacteriol. 195, 1152–1166. doi:10.1128/JB.01032-12

Lepp, D., Roxas, B., Parreira, V.R., Marri, P.R., Rosey, E.L., Gong, J., Songer, J.G., Vedantam, G., Prescott, J.F., 2010. Identification of novel pathogenicity loci in *Clostridium perfringens* strains that cause Avian necrotic enteritis. PLoS One 5. doi:10.1371/journal.pone.0010795

Letunic, I., Bork, P., 2011. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Res. 39, W475–W478. doi:10.1093/nar/gkr201

Li, L., Thøfner, I., Christensen, J.P., Ronco, T., Pedersen, K., Olsen, R.H., Li, L., Thøfner, I., Christensen, J.P., Ronco, T., Pedersen, K., 2017. Evaluation of the efficacy of an autogenous *Escherichia coli* vaccine in broiler breeders. Avian Pathol. 9457, 200–308. doi:10.1080/03079457.2016.1267857

Lina, G., Piémont, Y., Godail-gamot, F., Peter, M., 1999. Involvement of Panton-valentine leukocidin – producing *Staphylococcus aureus* in primary skin infections and pneumonia. Clin. inectious Dis. 1128–1132.

Loman, N.J., Constantinidou, C., Chan, J.Z.M., Halachev, M., Sergeant, M., Penn, C.W., Robinson, E.R., Pallen, M.J., 2012a. High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. Nat. Rev. Microbiol.

Loman, N.J., Misra, R. V, Dallman, T.J., Constantinidou, C., Gharbia, S.E., Wain, J., Pallen, M.J., 2012b. Performance comparison of benchtop high-throughput sequencing platforms. Nat. Biotechnol. 30, 562–562. doi:10.1038/nbt0612-562f

Lovland, A., Kaldhusdal, M., 2001. Severely impaired production performance in broiler flocks with high incidence of *Clostridium perfringens*-associated hepatitis. Avian Pathol. 30, 73–81.

Lyhs, U., Perko-Makela, P., Kallio, H., Brockmann, A., Heinikainen, S., Tuuri, H., Pedersen, K., 2013. Characterization of *Clostridium perfringens* isolates from healthy turkeys and from turkeys with necrotic enteritis. Poult. Sci. 92, 1750–1757. doi:10.3382/ps.2012-02903

Machha, V.R., Tischer, A., Moon-Tasson, L., Auton, M., 2016. The Von Willebrand Factor A1–Collagen III Interaction Is Independent of Conformation and Type 2 Von Willebrand Disease Phenotype. J. Mol. Biol. 429, 32–47. doi:10.1016/j.jmb.2016.11.014

Mahmmod, Y.S., Klaas, I.C., Enevoldsen, C., 2017. DNA carryover in milk samples from routine milk recording used for PCR-based diagnosis of bovine *Staphylococcus aureus* mastitis. J. Dairy Sci. 100, 5709–5716. doi:10.3168/jds.2016-12330

Maiden, M.C.J., Byrgraves, J.A., Feil, E., Morelli, G., Russell, J.E., Al, E., 1998. Multilocus sequence typing : A portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. 95, 3140–3145.

McDonel, J.L., 1980. *Clostridium perfringens* toxins (type a, b, c, d, e). Pharmac. Ther. 10, 617–655.

Mo, S.S., Kristoffersen, A.B., Sunde, M., Nødtvedt, A., Norström, M., 2016. Risk factors for

occurrence of cephalosporin-resistant *Escherichia coli* in Norwegian broiler flocks. Prev. Vet. Med. 130, 112–118. doi:10.1016/j.prevetmed.2016.06.011

Moore, R.J., 2016. Necrotic enteritis predisposing factors in broiler chickens. Avian Pathol. 9457, 1–22. doi:10.1080/03079457.2016.1150587

Mora, A., López, C., Herrera, A., Viso, S., Mamani, R., Dhabi, G., Alonso, M.P., Blanco, M., Blanco, J.E., Blanco, J., 2012. Emerging avian pathogenic *Escherichia coli* strains belonging to clonal groups O111: H4-D-ST2085 and O111: H4-D-ST117 with high virulence-gene content and zoonotic potential. Vet. Microbiol. 156, 347–352. doi:10.1016/j.vetmic.2011.10.033

Moulin-Schouleur, M., Reperant, M., Laurent, S., Bree, A., Mignon-Grasteau, S., Germon, P., Rasschaert, D., Schouler, C., 2007. Extraintestinal pathogenic *Escherichia coli* strains of avian and human origin: Link between phylogenetic relationships and common virulence patterns. J. Clin. Microbiol. 45, 3366–3376. doi:10.1128/JCM.00037-07

Murchan, S., Kaufmann, M.E., Deplano, A., Ryck, R. De, Struelens, M., Zinn, C.E., Fussing, V., Salmenlinna, S., Vuopio-varkila, J., Solh, E., Cuny, C., Witte, W., Tassios, P.T., Legakis, N., Leeuwen, W. Van, Belkum, A. Van, Vindel, A., Laconcha, I., Garaizar, J., Haeggman, S., Olsson-liljequist, B., Ransjo, U., 2003. Harmonization of Pulsed-Field Gel Electrophoresis Protocols for Epidemiological Typing of Strains of Methicillin-Resistant *Staphylococcus aureus* : a Single Approach Developed by Consensus in 10 European Laboratories and Its Application for Tracing the Spr. J. Clin. Microbiol. 41, 1574–1585. doi:10.1128/JCM.41.4.1574

Murray, N.E., 2000. Type I Restriction Systems: Sophisticated Molecular Machines (a Legacy of Bertani and Weigle). Microbiol. Mol. Biol. Rev. 64, 412–434. doi:10.1128/MMBR.64.2.412-434.2000

Nauerby, B., Pedersen, K., Madsen, M., 2003. Analysis by pulsed-field gel electrophoresis of the genetic diversity among *Clostridium perfringens* isolates from chickens. Vet. Microbiol. 94, 257–266.

Nilsson, O., Borjesson, S., Landen, A., Bengtsson, B., 2014. Vertical transmission of *Escherichia coli* carrying plasmid-mediated AmpC (pAmpC) through the broiler production pyramid. J. Antimicrob. Chemother. 69, 1497–1500. doi:10.1093/jac/dku030

Ohtani, K., Kawsar, H.I., Okumura, K., Hayashi, H., Shimizu, T., 2003. The VirR/VirS regulatory cascade affects transcription of plasmid-encoded putative virulence genes in

Clostridium perfringens strain 13. FEMS Microbiol. Lett. 222, 137–141. doi:10.1016/S0378-1097(03)00255-6

Okumura, K., Ohtani, K., Hayashi, H., Shimizu, T., 2008. Characterization of genes regulated directly by the VirR/VirS system in *clostridium perfringens*. J. Bacteriol. 190, 7719–7727. doi:JB.01573-07 [pii] 10.1128/JB.01573-07

Olsen, R.H., Stockholm, N.M., Permin, A., Christensen, J.P., Christensen, H., Bisgaard, M., 2011. Multi-locus sequence typing and plasmid profile characterization of avian pathogenic *Escherichia coli* associated with increased mortality in free-range layer flocks. Avian Pathol. 40, 437–444. doi:10.1080/03079457.2011.592822

Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T.G., Fookes, M., Falush, D., Keane, J.A., Parkhill, J., 2015. Roary: Rapid large-scale prokaryote pan genome analysis. Bioinformatics 31, 3691–3693. doi:10.1093/bioinformatics/btv421

Pires-dos-Santos, T., Bisgaard, M., Christensen, H., 2013. Genetic diversity and virulence profiles of Escherichia coli causing salpingitis and peritonitis in broiler breeders. Vet. Microbiol. 162, 873–880. doi:10.1016/j.vetmic.2012.11.008

Price, L.B., Johnson, J.R., Aziz, M., Clabots, C., Johnson, B., Tchesnokova, V., Nordstrom, L., et, al, 2013. The Epidemic of Extended-Spectrum-beta-Lactamase-Producing *Escherichia coli* ST131 Is Driven by a Single Highly Pathogenic. MBio 4, 1–10. doi:10.1128/mBio.00377-13.Editor

Price, L.B., Stegger, M., Hasman, H., Aziz, M., Larsen, J., Andersen, S., Pearson, T., Laurent, F., Keim, P., Skov, R., Aarestrup, F.M., 2012. *Staphylococcus aureus* CC398 : Host Adaptation and Emergence of Methicillin Resistance in Livestock. MBio 3(1), e00305-11.

Roberts, I.S., Road, O., 1999. Structure , assembly and regulation of expression of capsules in *Escherichia coli*. Mol. Microbiol. 31, 1307–1319.

Roe, C., Smith, D.E., Williamson, C.H.D., Aziz, M., Keim, P., Hepp, C.M., Driebe, E.M., Lemmer, D., Travis, J., Hicks, N.D., Schupp, J.M., Wagner, D.M., Engelthaler, D.M., Gillece, J.D., Sahl, J.W., Drees, K.P., 2016. NASP: an accurate, rapid method for the identification of SNPs in WGS datasets that supports flexible input and output formats. Microb. Genomics 2, 1–12. doi:10.1099/mgen.0.000074

Ronco, T., Stegger, M., Ng, K.L., Lilje, B., Lyhs, U., Andersen, P.S., 2017a. Genome analysis of *Clostridium perfringens* isolates from healthy and necrotic enteritis infected chickens

and turkeys. BMC Res. Notes 10, 1–6. doi:10.1186/s13104-017-2594-9

Ronco, T., Stegger, M., Olsen, R.H., Sekse, C., Nordstoga, A.B., Pohjanvirta, T., Lilje, B., Lyhs, U., Andersen, P.S., Pedersen, K., 2017b. Spread of avian pathogenic *Escherichia coli* ST117 O78 : H4 in Nordic broiler production. BMC Genomics 18, 1–8. doi:10.1186/s12864-016-3415-6

Saita, M., Bano, L., Gallazzi, D.D., 2010. Pathogenicity markers of Clostridium spp. in commercial turkeys. Ital.J.Anim.Sci. 8, 781–784. doi:10.4081/ijas.2009.781

Sakwinska, O., Giddey, M., Moreillon, M., Morisset, D., Waldvogel, A., Moreillon, P., 2011. *Staphylococcus aureus* host range and human-bovine host shift. Appl. Environ. Microbiol. 77, 5908–5915. doi:10.1128/AEM.00238-11

Santos Costa, S., Viveiros, M., Rosato, A.E., Melo-Cristino, J., Couto, I., 2015. Impact of efflux in the development of multidrug resistance phenotypes in *Staphylococcus aureus*. BMC Microbiol. 15, 232–247. doi:10.1186/s12866-015-0572-8

Seemann, T., 2014. Prokka: Rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069. doi:10.1093/bioinformatics/btu153

Selander, R.K., Caugant, D.A., Ochman, H., Musser, J.M., Gilmour, M.N., Whittam, T.S., 1986. Methods of Multilocus Enzyme Electrophoresis for Bacterial Population Genetics and Systematics. Appl. Environ. Microbiol. 51, 873–884.

Smith, E.M., Green, L.E., Medley, G.F., Bird, H.E., Fox, L.K., Schukken, Y.H., Kruze, J. V, Bradley, a J., Zadoks, R.N., Dowson, C.G., 2005. Multilocus sequence typing of intercontinental bovine *Staphylococcus aureus* isolates. J. Clin. Microbiol. 43, 4737–4743. doi:10.1128/JCM.43.9.4737

Songer, J.G., 1996. Clostridial enteric diseases of domestic animals. Clin. Microbiol. Rev. 9, 216–34.

Songer, J.G., Meer, R.R., 1996. Genotyping of *Clostridium perfringens* by polymerase chain reaction is a useful adjunct to diagnosis of Clostridial enteric diseases in animals. Anaerobe 2, 197–203. doi:DOI: 10.1006/anae.1996.0027

Stegger, M., Aziz, M., Chroboczek, T., Price, L.B., Ronco, T., Kiil, K., Skov, R.L., Laurent, F., Andersen, P.S., 2013. Genome analysis of *Staphylococcus aureus* ST291, a double locus variant of ST398, reveals a distinct genetic lineage. PLoS One 8(5).

Tagini, F., Greub, G., 2017. Bacterial genome sequencing in clinical microbiology : a pathogen-oriented review. Eur J Clin Microbiol Infect Dis. doi:10.1007/s10096-017-

3024-6

Umeda, K., Nakamura, H., Yamamoto, K., Nishina, N., Yasufuku, K., Hirai, Y., Hirayama, T., Goto, K., Hase, A., Ogasawara, J., 2017. Molecular and epidemiological characterization of staphylococcal foodborne outbreak of *Staphylococcus aureus* harboring *seg* , *sei* , *sem* , *sen* , *seo* , and *selu* genes without production of classical enterotoxi. Int. J. Food Microbiol. 256, 30–35. doi:10.1016/j.ijfoodmicro.2017.05.023

Vandendriessche, S., Vanderhaeghen, W., Larsen, J., de Mendonça, R., Hallin, M., Butaye, P., Hermans, K., Haesebrouck, F., Denis, O., 2014. High genetic diversity of methicillin-susceptible *Staphylococcus aureus* (MSSA) from humans and animals on livestock farms and presence of SCCmec remnant DNA in MSSA CC398. J. Antimicrob. Chemother. 69, 355–362. doi:10.1093/jac/dkt366

Vidal, J.E., Ma, M., Saputo, J., Garcia, J., Uzal, F.A., McClane, B.A., 2012. Evidence that the Agr-like quorum sensing system regulates the toxin production, cytotoxicity and pathogenicity of *Clostridium perfringens* type C isolate CN3685. Mol. Microbiol. 83, 179–194. doi:10.1111/j.1365-2958.2011.07925.x

Wade, B., Keyburn, A.L., 2015. The true cost of necrotic enteritis. World Poult. 31, 16–17.

Wade, B., Keyburn, A.L., Seemann, T., Rood, J.I., Moore, R.J., 2015. Binding of *Clostridium perfringens* to collagen correlates with the ability to cause necrotic enteritis in chickens. Vet. Microbiol. 180, 299–303. doi:10.1016/j.vetmic.2015.09.019

Xu, J., Tan, X., Zhang, X., Xia, X., Sun, H., 2015. The diversities of staphylococcal species, virulence and antibiotic resistance genes in the subclinical mastitis milk from a single Chinese cow herd. Microb. Pathog. 88, 29–38. doi:10.1016/j.micpath.2015.08.004

Zadoks, R.N., Middleton, J.R., McDougall, S., Katholm, J., Schukken, Y.H., 2011. Molecular epidemiology of mastitis pathogens of dairy cattle and comparative relevance to humans. J. Mammary Gland Biol. Neoplasia 16, 357–372. doi:10.1007/s10911-011-9236-y

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., Aarestrup, F.M., Larsen, M.V., 2012. Identification of acquired antimicrobial resistance genes. J. Antimicrob. Chemother. 67, 2640–2644. doi:10.1093/jac/dks261

Zankari, E., Hasman, H., Kaas, R.S., Seyfarth, A.M., Agersø, Y., Lund, O., Larsen, M.V., Aarestrup, F.M., 2013. Genotyping using whole-genome sequencing is a realistic alternative to surveillance based on phenotypic antimicrobial susceptibility testing 771–777. doi:10.1093/jac/dks496

Zurfluh, K., Wang, J., Klumpp, J., Nüesch-Inderbinen, M., Fanning, S., Stephan, R., 2014. Vertical transmission of highly similar blaCTX-M-1-harboring IncI1 plasmids in *Escherichia coli* with different MLST types in the poultry production pyramid. Front. Microbiol. 5, 1–7. doi:10.3389/fmicb.2014.00519

Aarestrup, Wegener, H.C., Rosdahl, V.T., 1995a. Evaluation of phenotypic and genotypic methods for epidemiological typing of *Staphylococcus aureus* isolates from bovine mastitis in Denmark. Vet. Microbiol. 45, 139–150. doi:10.1016/0378-1135(95)00043-A

Aarestrup, Andersen, J.K., Jensen, N.E., 1995b. Lack of Staphylococcal enterotoxin production among strains of *Staphylococcus aureus* from bovine mastitis in Denmark. Acta Vet Scand 36, 273–275.

Appendix I

Paper I

Ronco T, Stegger M, Andersen PS, Pedersen K, Li L, Thøfner ICN, Olsen RH. 2016. Draft genome sequences of two avian pathogenic *Escherichia coli* strains of clinical importance, E44 and E51. Genome Announc. 4, e00768-16.

AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements

Draft Genome Sequences of Two Avian Pathogenic *Escherichia coli* Strains of Clinical Importance, E44 and E51

Troels Ronco,[a] Marc Stegger,[b] Paal S. Andersen,[b] Karl Pedersen,[a] Lili Li,[c] Ida C. N. Thøfner,[d] Rikke H. Olsen[d]

National Veterinary Institute, Technical University of Denmark, Frederiksberg, Denmark[a]; Department of Microbiology and Infection Control, Statens Serum Institut, Copenhagen, Denmark[b]; College of Light Industry and Food Sciences, South China University of Technology, Guangzhou, Guangdong, People's Republic of China[c]; Department of Veterinary Disease Biology, University of Copenhagen, Frederiksberg, Denmark[d]

Avian pathogenic *Escherichia coli* strains have remarkable impacts on animal welfare and the production economy in the poultry industry worldwide. Here, we present the draft genomes of two isolates from chickens (E44 and E51) obtained from field outbreaks and subsequently investigated for their potential for use in autogenous vaccines for broiler breeders.

Received 8 June 2016 **Accepted** 10 June 2016 **Published** 4 August 2016

Citation Ronco T, Stegger M, Andersen PS, Pedersen K, Li L, Thøfner ICN, Olsen RH. 2016. Draft genome sequences of two avian pathogenic *Escherichia coli* strains of clinical importance, E44 and E51. Genome Announc 4(4):e00768-16. doi:10.1128/genomeA.00768-16.

Address correspondence to Troels Ronco, troro@vet.dtu.dk.

Avian pathogenic *Escherichia coli* (APEC) causing colibacillosis in commercial poultry is an important bacterial pathogen (1). Whereas "colibacillosis" commonly refers to systemic or localized infection in broilers, ascending infections due to *E. coli* in breeders and layers may lead to infection of the reproductive tract (2), with significant impacts on animal welfare and the poultry production economy (3). Nevertheless, there are few commercially available vaccines for the protection of layers and broilers against *E. coli* infection. Consequently, the use of autogenous *E. coli* vaccines is a common practice (4). The aims of using these vaccines are two-fold: direct protection of the breeders and indirect protection of the offspring through the passage of maternally derived antibodies.

In recent years, outbreaks due to *E. coli* in broiler breeders and broilers have increased in Scandinavian countries, expediting the introduction of a new autogenous *E. coli* vaccine program for broiler breeders. Here we present the draft genomes of two *E. coli* isolates (E44 and E51) included in this autogenous vaccine.

Fragment libraries were constructed using a Nextera XT kit (Illumina) followed by 251-bp paired-end sequencing (MiSeq; Illumina) according to manufacturer's instructions. Genomics Workbench 6.5 (CLC bio) was used for *de novo* assembly of the raw reads. It resulted in totals for size of assembly/N_{50} of 5,125,126 bp/83,776 bp and 5,178,940 bp/100,046 bp, total numbers of contigs of 195 and 217, and average coverages/G+C contents of 91×/50.5% and 58×/50.5% for E44 and E51, respectively.

The contigs were annotated in the NCBI Prokaryotic Genome Automatic Annotation Pipeline (PGAAP) (5). In total, E44 had 5,171 putative genes, of which 4,868 were protein-coding sequences (CDSs), whereas E51 had 5,305 putative genes, including 4,986 protein CDSs.

Various types of virulence genes that previously have been associated with APEC isolates (6) were extracted from NCBI and identified in the draft genomes using MyDbFinder 1.1 (https://cge.cbs.dtu.dk/services/MyDbFinder/). E44 carried fewer virulence genes (*fimA*, *fimC*, *iroN*, *iss*, *iucA*, *iucD*, *ompA*, and *vat*) than E51 (*cvaB/C*, *cvi*, *fimA*, *fimC*, *fyuA*, *ibeA*, *iroN*, *irp2*, *iss iucA*, *iucD*, and *ompA*). According to PathogenFinder (7), both E44 and E51 were predicted to be human pathogens, with probabilities of 93% and 94%, because they matched 533 and 856 pathogenic families, respectively. None of the strains carried any antibiotic resistance genes, as verified using ResFinder 2.1 (8). *In silico* typing using MLST 1.8 (9) and SerotypeFinder 1.1 (10) showed that the sequence types (STs)/serotypes of E44 and E51 were O78:H4/ST117 and O2:H5/ST140, respectively. Field production data from farms using the E44/E51-based vaccine, experimental data obtained from *in vivo* infection models, and further genome analyses could provide useful knowledge regarding development of new vaccines and insight into virulent properties.

Accession number(s). The two whole-genome shotgun projects have been deposited in DDBJ/ENA/GenBank under the accession numbers LXWV00000000 (E44) and LYPJ00000000 (E51). The versions described in this paper are the first versions.

REFERENCES

1. **Dho-moulin M, Fairbrother JM.** 1999. Avian pathogenic *Escherichia coli* (APEC). Vet Res **30:**299–316.
2. **Olsen RH, Thøfner IC, Pors SE, Pires dos Santos T, Christensen JP.** 2016. Experimental induced avian *E. coli* salpingitis: significant impact of strain and host factors on the clinical and pathological outcome. Vet Microbiol **188:**59–66. http://dx.doi.org/10.1016/j.vetmic.2016.04.011.
3. **Landman WJ, van Eck JH.** 2015. The incidence and economic impact of the *Escherichia coli* peritonitis syndrome in Dutch poultry farming. Avian Pathol **44:**370–378. http://dx.doi.org/10.1080/03079457.2015.1060584.
4. **Landman WJ, Buter GJ, Dijkman R, van Eck JH.** 2014. Molecular typing of avian pathogenic *Escherichia coli* colonies originating from outbreaks of *E. coli* peritonitis syndrome in chicken flocks. Avian Pathol **43:**345–356. http://dx.doi.org/10.1080/03079457.2014.935291.
5. **Angiuoli SV, Gussman A, Klimke W, Cochrane G, Field D, Garrity G, Kodira CD, Kyrpides N, Madupu R, Markowitz V, Tatusova T, Thomson N, White O.** 2008. Toward an online repository of standard operating procedures (SOPs) for (meta)genomic annotation. Omics **12:**137–141. http://dx.doi.org/10.1089/omi.2008.0017.
6. **Ewers C, Li G, Wilking H, Kießling S, Alt K, Antáo EM, Laturnus C, Diehl I, Glodde S, Homeier T, Böhnke U, Steinrück H, Philipp HC,**

Wieler LH. 2007. Avian pathogenic, uropathogenic, and newborn meningitis-causing *Escherichia coli*: how closely related are they? Int J Med Microbiol **297:**163–176. http://dx.doi.org/10.1016/j.ijmm.2007.01.003.

7. **Cosentino S, Voldby Larsen M, Møller Aarestrup F, Lund O**. 2013. Pathogenfinder—distinguishing friend from foe using bacterial whole genome sequence data. PLoS One **8:**e77302. http://dx.doi.org/10.1371/journal.pone.0077302.
8. **Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV**. 2012. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother **67:**2640–2644. http://dx.doi.org/10.1093/jac/dks261.
9. **Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Jelsbak L, Sicheritz-Pontén T, Ussery DW, Aarestrup FM, Lund O**. 2012. Multilocus sequence typing of total-genome-sequenced bacteria. J Clin Microbiol **50:**1355–1361. http://dx.doi.org/10.1128/JCM.06094-11.
10. **Joensen KG, Tetzschner AM, Iguchi A, Aarestrup FM, Scheutz F**. 2015. Rapid and easy *in silico* serotyping of *Escherichia coli* using whole-genome sequencing (WGS) data. J Clin Microbiol **53:**2410–2426. http://dx.doi.org/10.1128/JCM.00008 –15.

Appendix II

Paper II

Ronco T, Stegger M, Olsen RH, Sekse C, Nordstoga AB, Pohjanvirta T, Lilje B, Lyhs U, Andersen PS, Pedersen K. 2017. Spread of avian pathogenic *Escherichia coli* ST117 O78:H4 in Nordic broiler production. BMC Genomics 18, 1–8.

Ronco et al. BMC Genomics (2017) 18:13
DOI 10.1186/s12864-016-3415-6

BMC Genomics

RESEARCH ARTICLE Open Access

Spread of avian pathogenic *Escherichia coli* ST117 O78:H4 in Nordic broiler production

Troels Ronco[1*], Marc Stegger[2], Rikke Heidemann Olsen[3], Camilla Sekse[4], Anne Bang Nordstoga[4], Tarja Pohjanvirta[5], Berit Lilje[2], Ulrike Lyhs[1], Paal Skytt Andersen[2] and Karl Pedersen[1]

Abstract

Background: *Escherichia coli* infections known as colibacillosis constitute a considerable challenge to poultry farmers worldwide, in terms of decreased animal welfare and production economy. Colibacillosis is caused by avian pathogenic *E. coli* (APEC). APEC strains are extraintestinal pathogenic *E. coli* and have in general been characterized as being a genetically diverse population. In the Nordic countries, poultry farmers depend on import of Swedish broiler breeders which are part of a breeding pyramid. During 2014 to 2016, an increased occurrence of colibacillosis on Nordic broiler chicken farms was reported. The aim of this study was to investigate the genetic diversity among *E. coli* isolates collected on poultry farms with colibacillosis issues, using whole genome sequencing.

Methods: Hundred and fourteen bacterial isolates from both broilers and broiler breeders were whole genome sequenced. The majority of isolates were collected from poultry with colibacillosis on Nordic farms. Subsequently, comparative genomic analyses were carried out. This included *in silico* typing (sero- and multi-locus sequence typing), identification of virulence and resistance genes and phylogenetic analyses based on single nucleotide polymorphisms.

Results: In general, the characterized poultry isolates constituted a genetically diverse population. However, the phylogenetic analyses revealed a major clade of 47 closely related ST117 O78:H4 isolates. The isolates in this clade were collected from broiler chickens and breeders with colibacillosis in multiple Nordic countries. They clustered together with a human ST117 isolate and all carried virulence genes that previously have been associated with human uropathogenic *E. coli*.

Conclusions: The investigation revealed a lineage of ST117 O78:H4 isolates collected in different Nordic countries from diseased broilers and breeders. The data indicate that the closely related ST117 O78:H4 strains have been transferred vertically through the broiler breeding pyramid into distantly located farms across the Nordic countries.

Keywords: APEC, Colibacillosis, Comparative genomics, Phylogenetic analysis, Virulence factors

Background

Escherichia coli infections in poultry constitute a severe animal health issue and a considerable burden to farmers worldwide, in terms of decreased animal welfare and production economy [1, 2]. Disease in poultry caused by avian pathogenic *E. coli* (APEC) may cause a wide range of extraintestinal symptoms, collectively termed colibacillosis. APEC belong to the group of extraintestinal pathogenic *E. coli* (ExPEC) [3–5], that also includes the pathotypes; uropathogenic *E. coli* (UPEC), neonatal-meningitis *E. coli* (NMEC) and septicemic *E. coli*. All groups have been associated with disease in both humans and animals [3, 4], and it has been reported that human ExPEC strains are closely related to APEC strains, suggesting that poultry could constitute a reservoir of zoonotic APEC strains [3, 6, 7].

In general, APEC isolates from chickens constitute a genetically diverse population with numerous of different serogroups and sequence types (STs). The most commonly observed serogroups are O1, O2 and O78 [6–9], and multilocus sequence typing (MLST) has shown that STs 10, 48, 95 and 117 have been frequently observed [5, 10, 11]. Several types of virulence genes are commonly identified in APEC as well as in human

* Correspondence: troro@vet.dtu.dk
[1]National Veterinary Institute, Technical University of Denmark, Bülowsvej 27, 1870 Frederiksberg C, Denmark
Full list of author information is available at the end of the article

ExPEC [7, 12] (Additional file 1: Table S1). These are often carried on virulence plasmids and pathogenicity islands (PAIs) [13–15].

Nordic broiler production depends on a breeding pyramid where Swedish grandparents are mainly imported from Scotland and used for breeding of parents for export to farms in the rest of the Nordic countries [16, 17]. Interestingly, previous studies indicated that extended-spectrum beta-lactamase (ESBL)-producing *E. coli* can be transmitted vertically from parents to offspring through the breeding pyramid [17–19]. Hence, if great grandparents are infected with virulent *E. coli*, they can potentially disseminate vertically to Swedish grandparents and hereafter to parents and broilers across the Nordic countries.

From 2009 until late 2014, the mortality on Danish poultry farms has on average been decreasing. Hereafter, the mortality has increased to >4.0% in 2015 and it has been suggested that colibacillosis in both parents and broilers has played a significant role [20]. In the same period similar problems with colibacillosis and increased mortality have also been observed on Finnish and Norwegian farms (Magne Hansen, Animalia, pers. comm). The aim of this study was to investigate, using whole genome sequencing, the genetic diversity and potential relatedness of APEC isolates associated with increased mortality and colibacillosis in Nordic countries.

Methods

E. coli isolates

In this study, 107 bacterial isolates from Danish ($n = 74$), Finnish ($n = 15$), Norwegian ($n = 16$) and Polish ($n = 2$) farms were analyzed. Additionally, assembled draft genomes obtained from seven bacterial isolates collected from diseased Danish chickens, were kindly provided by the Danish poultry industry, and their isolation ID have in this study been assigned a capital "A" (Additional file 2: Table S2). In total, the 114 isolates were collected from 88 different farms and if isolates were from the same farm they were in general collected from different houses. The majority of isolates were collected from diseased broiler chickens and parents (layer hens) and diseased birds were diagnosed with a generalized *E. coli* infection, whereas 15 isolates were collected from healthy birds (Additional file 2: Table S2). The Danish isolates were collected from all parts of the country by the two commercial laboratories at LVK (Landbrugets Veterinære Konsulenttjeneste, Hobro, Denmark) and the poultry slaughterhouse, Danpo (Danpo A/S, Aars, Denmark). The Danish farms were not geographically clustered, but distributed evenly throughout the regions with poultry production. Notably, the draft genomes of two Danish isolates (E44 and E51) have previously been annotated and deposited in DDBJ/ENA/GenBank under the accession numbers LXWV00000000 (E44) and LYPJ00000000 (E51) [21], due to their inclusion in a Danish autogenous vaccine program. Finnish isolates were collected from the Southwestern part of Finland by the Finnish Food Safety Authority (Evira). In Norway, the isolates were collected from central, South-eastern and Western parts of the country by the Norwegian Veterinary Institute (NVI). In both Finland and Norway, the samples were collected from regions where most broiler farms are located.

The Illumina reads sequenced in this study were deposited in the NCBI SRA [22] under the study accession number SRP092633.

DNA purification and sequencing

Isolates were grown overnight at 37 °C on blood agar (Columbia agar base [Oxoid, Hampshire, UK]) supplemented with 5% calf blood [SSI, Copenhagen, DK]). Single colonies were harvested directly from the agar plates and genomic DNA was purified using the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions. The DNA libraries were generated using Nextera XT kit (Illumina Inc., San Diego, Ca) according to manufacturer's instructions. Finally, Illumina's MiSeq platform was used for paired-end DNA sequencing with a read length of 2 × 251 bp for all isolates except the seven isolates from the Danish poultry industry which were sequenced with a read length of 2 × 300 bp.

De novo assembly and typing

Raw reads were *de novo* assembled using CLC bio's Genomics Workbench (GW) v6.5 (Qiagen, Aarhus, Denmark) with default setting and a threshold on contigs of minimum 500 nt. Subsequently, the *de novo* assembled contigs were MLST [23] and serotyped [24] *in silico* using online typing tools [25].

Core genome diversity

The genetic relationship between all 114 isolates was investigated using single nucleotide polymorphisms (SNPs). SNPs were identified using NASP 1.0 [26] by aligning Illumina reads against *E. coli* strain CFT073 (GenBank accession no. AE014075), using the Burrows-Wheeler Aligner (BWA) [27] after removal of duplicated regions in the reference using NUCmer [28]. GATK Unified Genotyper [29] was used to identify variant positions and to remove positions with <90% unambiguous base calls, as well as SNPs in positions that did not meet a minimum coverage requirement of ≥10×. Subsequently, a phylogenetic tree model was constructed using the maximum-likelihood algorithm implemented in PhyML [30] using Smart Model Selection and the Bayesian Information Criterion with 100 bootstrap replicates and visualized using iTOL v3.1 [31].

Further investigations were performed on all identified ST117 isolates from this study ($n = 62$) together with all 21 identified ST117 *E. coli* strains available at [32] (Additional file 3: Table S3). SNPs were identified as described above but with purging of recombinant regions using Gubbins v1.4.4 with standard settings of five iterations. Phylogenetic analyses on the purged dataset were performed as previously described.

Identification of virulence and resistance genes

Various types of virulence genes were identified in *de novo* assembled contigs using MyDbFinder v1.1 [33] and antibiotic resistance genes were identified using ResFinder v2.1 [34]. Further descriptions and Genbank accession numbers of the selection of virulence genes are found in Additional file 1: Table S1. In some cases, CLC bio's Genomics workbench was used to verify the presence of the open reading frames by BLASTN and mapping of reference genes to the *de novo* assembled contigs.

Results

The majority of the isolates (66/114) were sequenced to an average coverage of ≥50 fold, whereas 35/114 of the draft genomes had an average coverage of >30. The rest had an average coverage of >18, whereas seven draft genomes provided by the Danish poultry industry exhibited slightly less coverage. Assembly metrics (average coverage, N50, number of contigs and assembly size) can be found in Additional file 4: Table S4.

Serotyping and MLST

Serotype genes were in the majority of the 114 assembled genomes, identified with thresholds of ≥90% nucleotide identity, ≥90% coverage of the query and a sequence depth of >10×. However, in eight isolates (E29, E50, E52, E56, E64, E66, E69 and E91) the O-type genes were on average identified with 67% coverage of the query and ≥90% nucleotide identity, whereas no O-type genes were identified in nine isolates (Additional file 2: Table S2). The isolates showed a high diversity with a total of 33 different serotypes, and MLST analyses identified 29 different STs. The most prevalent serotype was O78:H4 observed among 43% (49/114) of the isolates, whereas 54% (62/114) were found to be of ST117 (Additional file 2: Table S2). Of all isolates, 61% (70/114) could be divided into six groups associated with the same serotype and ST (Table 1). Notably, all six groups presented in Table 1 were also closely related according to the SNP analysis (Fig. 1).

Phylogenetic analyses

The SNP calling based on all 114 isolates had a total of 145,637 variant positions identified in ~49% of the reference genome. The phylogenetic analysis revealed a large clade of 62 isolates from both diseased broilers and parents collected in Denmark, Finland and Norway. Isolates from healthy chickens were not related to this clade (Fig. 1). All isolates from this clade belonged to ST117 and 47 of these had serotype O78:H4 whereas eight had O53:H5 (Fig. 1). It was not possible to identify O-type genes in six isolates from this clade, but they all carried the H4 gene. A single isolates (E24) was of serotype O161:H4 (Fig. 1) (Additional file 2: Table S2). Minor clusters of few isolates with identical serotype and ST were present but none of them contained isolates from more than two different countries (Fig. 1).

The SNP calling based on all 83 ST117 genomes from both this study and those obtainable from the public domain at NCBI [32] had a total of 13,215 variant positions identified in ~64% of the genomes, and with 2,617 SNPs remaining after purging of recombinant regions. The analysis identified a major clade primarily consisting of 47 O78:H4 isolates from both diseased broilers and parents collected on 34 different farms in Denmark, Finland and Norway (Fig. 2). On average, the length between these 47 isolates was 23 SNPs. In four isolates (E46, E53, E54, and E72) from this clade no O-type genes were identified, whereas only one isolate in this clade (E13) belonged to a serogroup different from O78 (Fig. 2). Furthermore, strain GN02004 obtained from the RefSeq archive at NCBI (Additional file 3: Table S3) belonged to serotype O24:H4. The 47 O78:H4 isolates differed by at least 50 SNPs from the O53:H4 clade (Fig. 2).

Identification of virulence and resistance genes

In general, different combinations of virulence genes were identified in the 114 draft genomes with thresholds of ≥90% nucleotide identity and ≥90% coverage of the query sequence. (Table 1 and Additional file 2: Table S2). All 47 ST117 O78:H4 isolates carried a unique combination of nine virulence genes compared to the eight ST117 O53:H4 isolates (Table 1). However, the isolates in this cluster did not carry any of the investigated virulence genes that were not also present in other isolates, i.e., no virulence genes were unique for this cluster. *hlyA* was not identified in any of the isolates whereas *fimA*/*C* and *ompA* were found in almost all isolates. *papC*, *tsh*, *vat* and *ibeA* were only associated with diseased poultry and *iroD*, *iroN* and *iss* were found in ≥94% (93/99) of the isolates from diseased birds (Table 1). The majority (79/114) of all isolates did not carry antibiotic resistance genes. Only few (6/47) of the ST117 O78:H4 isolates carried antibiotic resistance genes, whereas resistance genes were more common among the other groups (Additional

Table 1 Virulence gene content among 114 *E. coli* isolates

	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Healthy	Diseased
Isolates	47	8	5	4	3	3	15	99
Farms	34	8	4	4	2	3	15	73
Serotype	O78:H4	O53:H4	O103:H2	O18ac:H7	O149:H23	O5:H10	-	-
ST	117	117	1146	95	1163	93	-	-
fimA	+	+	+	+	+	1/3	12/15	94/99
fimC	+	+	+	+	+	+	13/15	96/99
papC								4/99
tsh				+		+		20/99
fyuA		+		3/4			1/15	28/99
iroD	+	+		+	+	+	4/15	94/99
iroN	+	+		+	+	+	3/15	93/99
irp2		+		+			1/15	30/99
iucA	+	+	+	+		+	7/15	84/99
iucD	+	+	+	+		1/3	7/15	84/99
cva		+	+	+	+	+	12/15	39/99
cvi		+	+	+	+	+	12/15	39/99
iss	+	+	2/5	+	+	+	7/15	97/99
ompA	+	+	+	+	+	+	15/15	99/99
vat	+	+		+				71/99
hlyA								
ibeA				+		1/3		13/99

The table shows the virulence gene content among *E coli* groups of identical ST and serotype. "+" indicates presence of genes. If a virulence gene was not present among all isolates in a group, the ratio of isolates that carried the gene is presented. The number of isolates and the number of different farms they were collected from, in each group is shown. The isolates from these groups were also closely related according to the SNP analysis, and the groups are highlighted in colors in Fig. 1. Furthermore, the ratios of isolates from healthy and diseased poultry that carried virulence genes are shown

file 2: Table S2). The most commonly identified resistance genes were against β-lactams, sulphonamides or streptomycin, and less often against tetracycline or trimethoprim.

Discussion

In recent years (2014–2016), an increase in cases of colibacillosis on Nordic poultry farms has caused a raise in mortality and economic losses [20] (Magne Hansen, Animalia, pers. comm). Therefore, the genetic diversity among 114 *E. coli* isolates mainly collected from diseased animals on poultry farms with colibacillosis were investigated. In agreement with previous studies [5, 7, 9, 11], it was found that the poultry isolates were a genetically diverse population. However, we identified a group of 47 closely related ST117 O78:H4 isolates collected from diseased broilers and parents in multiple Nordic countries, which shared a similar genetic background (Fig. 2). In concordance, it has previously been reported that APEC isolates are widely associated with serogroup O78 and ST117 [7, 11]. These 47 isolates were not related to any of the isolates from healthy chickens (Fig. 1). On average, the distance between these 47 isolates was 23 SNPs, whereas the distance to the closest related ST117 O53:H4 isolate outside this group was 50 SNPs (Fig. 2). Additionally, all 47 ST117 O78:H4 carried an identical and unique combination of virulence genes compared to all other investigated isolates (Table 1). Thus, according to both the investigation of virulence profiles and the SNP analyses the 47 ST117 O78:H4 isolates define a distinct lineage. The isolates from this lineage carried nine genes that encode virulence factors important in the pathogenesis of avian colibacillosis (Table 1). Interestingly, O78 strains have been suggested to be the main cause of avian colisepticemia together with O1 and O2 strains [35]. The O-antigen capsule allows bacteria to avoid the host's innate immune response and studies have shown that this LPS capsule is required during systemic infections [35, 36]. Type I fimbriae encoded by *fimA/C*, have been shown to be necessary for initial colonization of the respiratory system [37]. Additionally, previous studies suggest that *iucA/D* and *iroD* which encode aerobactin and salmochelin siderophores respectively, are specifically important for iron acquisition in the extraintestinal environment of chickens [38]. As in this study, the

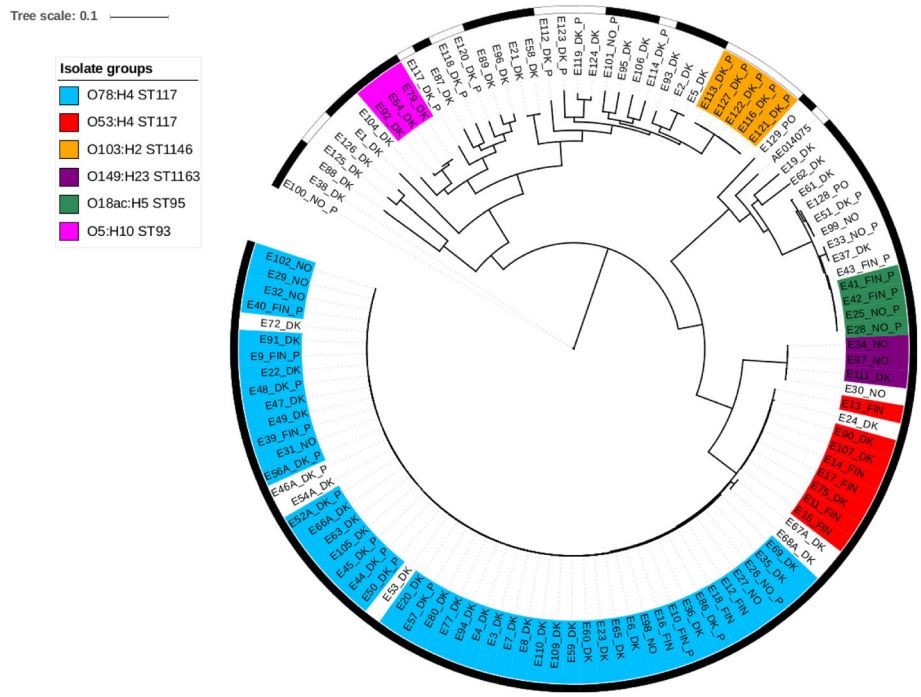


Fig. 1 Maximum-likelihood tree of 114 *E. coli* isolates based on 145,637 core SNPs. The analysis shows a clade of 62 APEC isolates collected from both broilers and parents (P) on Danish (DK), Finnish (FIN) and Norwegian (NO) chicken farms. All 62 isolates belonged to ST117. Isolate groups of the same serotype and ST are presented in identical colors. Isolates collected from diseased animals are marked with a black strip, whereas the white strip indicates isolates from healthy chickens. *E. coli* strain CFT073 served as reference and the scale indicates substitutions per site

virulence factor increased serum survival encoded by *iss,* was previously identified in APEC strains [7, 12] as well as the vacuolating autotransporter toxin Vat, but their exact role in the pathogenesis needs to be further elucidated [15, 39]. It should also be noticed that some of the virulence genes (*fimA*/*C*, *ompA*) were found among almost all analyzed isolates, both from healthy and diseased birds, which indicates that they are not only involved in avian colibacillosis. Additionally, a wide range of other virulence factors have also been suggested to be associated with APEC [40]. None of the investigated virulence genes (Table 1) were found exclusively among the O78:H4 cluster and it may therefore be suggested that other, yet undefined virulence mechanisms were partly responsible for the high virulence of this strain. Thus, a comparative study to reveal more precisely why the O78:H4 ST117 lineage was considerably associated with increased mortality among Nordic broilers and breeders from 2014–2016 could be interesting to carry out.

In the Nordic countries, all poultry farms receive their parents from Swedish hatcheries where grandparents are imported from Scotland [16, 17]. It was not possible to verify how many Swedish parent flocks that the birds included in this study originated from, which could have further revealed the extent of the colibacillosis issues observed on Nordic poultry farms. Finding highly similar isolates in broilers and parents from distantly located farms that share one common source for parent animals strongly support a vertical dissemination of ST117 O78:H4 isolates from grandparents and great grandparent. Vertical transmission of pathogenic *E. coli* has previously been observed in other studies [17–19]. Unfortunately, *E. coli* isolates from diseased grandparents or great grandparents were not available to further confirm this hypothesis. It could have been interesting to investigate samples from parents and their corresponding offspring but a parent flock can be origin to several different broiler flocks. Thus, it is not possible to collect samples directly from parent/offspring pairs [16]. Instead, an in vivo infection study of parents and their offspring could be carried out.

The 47 O78:H4 ST117 isolates carried various types of virulence genes (Table 1) that previously have been identified in both APEC and human UPEC isolates. Additionally, they were closely related to the ST117 O24:H4 *E. coli* strain GN02004 from NCBI (Fig. 2), which previously has been collected from human body fluids. (Additional file 3: Table S3). Therefore, it could be speculated whether the origin of these colibacillosis cases could have been the introduction of human UPEC to grandparents or great grandparents in the upper parts of

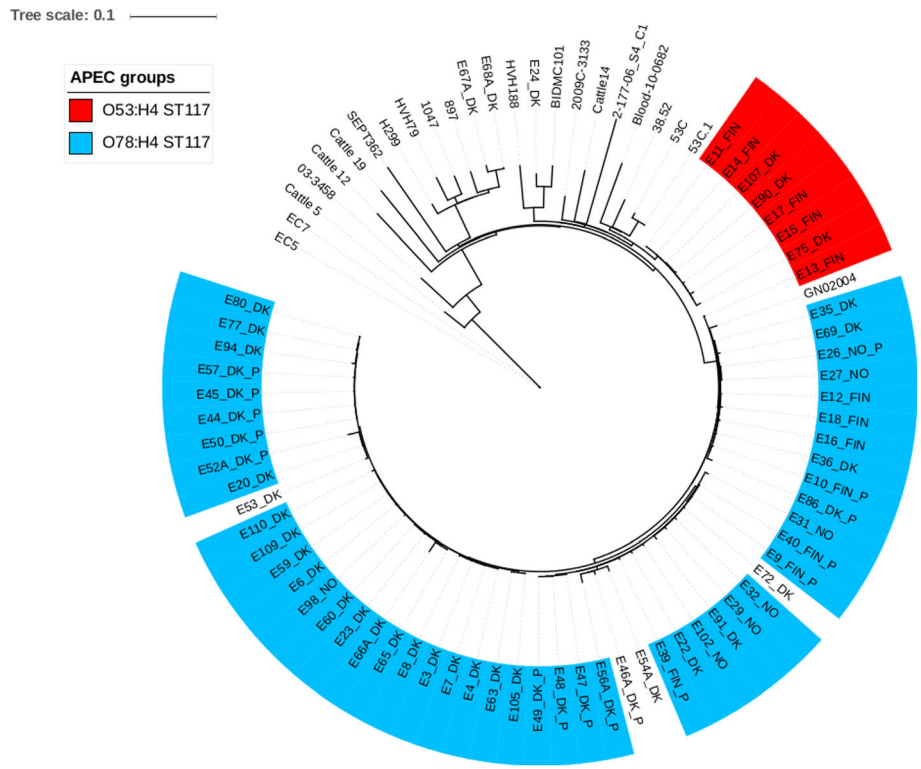


Fig. 2 Maximum likelihood tree of 83 ST117 *E. coli* based on 2,617 purged core SNPs. The analysis includes 62 ST117 poultry isolates and 21 international strains. A distinct clade of 53 closely related APEC isolates collected from both broilers and parents (P) on Danish (DK), Finnish (FIN) and Norwegian (NO) chicken farms was identified. Fourty-seven of these isolates were of serotype O78:H4 (blue), whereas a single was O53:H4 (red). All isolates from this study were from diseased birds and the reference strain *E. coli* strain CFT073 is not included. The scale indicates substitutions per site

the breeding pyramid. Strain GN02004 was sampled in USA but we could not obtain further information regarding possible dissemination via travel activities to Europe. Thus, diseased chickens could potentially constitute a zoonotic risk to e.g., farmers or other people who get in contact with them [3, 7]. Additionally, *E. coli* strains have been found to end up in the broiler chicken meat [17].

Since the increase in the occurrence of colibacillosis has a considerable impact on animal welfare and production economy, it is of great importance to obtain more in-depth knowledge regarding APEC and colibacillosis and to develop vaccines that possibly could provide immunization of the poultry. Interestingly, *E. coli* strain E44 [21] collected from a diseased parent bird, was here shown to be a part of the major ST117 O78:H4 linage. In 2015, strain E44 was selected for a Danish autovaccine program due to its suggested relation to the increase in colibacillosis on Danish poultry farms. However, the efficiency of the program remains to be evaluated.

Conclusion

Genomic investigation of APEC isolates collected from diseased chickens on Nordic poultry farms revealed the presence of a predominant lineage of ST117 O78:H4 isolates. The analyses indicated that the ST117 O78:H4 strains have been transmitted vertically through the broiler breeding pyramid and contributed considerably to the increase of colibacillosis cases observed on Nordic poultry farms from 2014 to 2016.

Additional files

Additional file 1: Table S1. APEC- and human UPEC-associated virulence genes. The pdf file contains descriptions of APEC- and human UPEC-associated virulence genes including GenBank accessions numbers and references. (PDF 295 kb)

Additional file 2: Table S2. *E. coli* isolates from chicken farms. The pdf file contains information regarding the farm, house, source, country sample date and state of animal associated with each of the 114 poultry isolates. Furthermore, the serotypes and STs are presented together with virulence and resistance gene content. (PDF 806 kb)

Additional file 3: Table S3. ST117 *E. coli* strains from NCBI. The pdf file contains information regarding all 21 international ST117 *E. coli* isolates available at NCBI [32]. The host, collection date, country and isolation source associated with each isolate are included together with the GenBank accession number. (PDF 295 kb)

Additional file 4: Table S4. Assembly metrics for *E. coli* isolates. This pdf file shows assembly metrics (average coverage, N50, number of contigs and assembly size) for the 114 isolates. (PDF 436 kb)

Acknowledgements
We are very thankful for the support from the Danish Poultry Levy Fond and we will like to thank Elvira Chapka and Mette Holm from Statens Serum Institut for excellent DNA sequencing work.

Funding
This investigation was supported by a grant from the Danish Poultry Levy Fond.

Availability of data and material
The clinical *E. coli* isolates and the whole genome sequences generated during and/or analyzed during the current study will be available from the corresponding author on reasonable request.

Authors' contributions
ABN, BL, CS, KP, MS, RHO, TP, TR and UL performed the experiments designed by KP, MS and RHO and TR. KP, MS, PSA, RHO and TR performed data analyses and interpretations. TR wrote the manuscript which finally was edited and approved by all authors.

Competing interests
The authors declare that they have no competing interests in this section.

Consent for publication
Not applicable.

Ethics approval and consent to participate
Not applicable.

Author details
[1]National Veterinary Institute, Technical University of Denmark, Bülowsvej 27, 1870 Frederiksberg C, Denmark. [2]Statens Serum Institut, Department of Microbiology and Infection Control, Artillerivej 5, 2300 Copenhagen S, Denmark. [3]Department of Veterinary Disease Biology, University of Copenhagen, Stigbøjlen 4, 1870 Frederiksberg C, Denmark. [4]Norwegian Veterinary Institute, Ullevaalsveien 68, 0454 Oslo, Norway. [5]Finnish Food Safety Authority, Veterinary Bacteriology, Neulaniementie 4, FI-70210 Kuopio, Finland.

Received: 14 September 2016 Accepted: 12 December 2016
Published online: 03 January 2017

References

1. Landman WJ, van Eck JH. The incidence and economic impact of the *Escherichia coli* peritonitis syndrome in Dutch poultry farming. Avian Pathol. 2015;9457:1–27.
2. Zhuang Q-Y, Wang S-C, Li J-P, Liu D, Liu S, Jiang W-M, et al. A clinical survey of common avian infectious diseases in China. Avian Dis. 2014;58:297–302.
3. Moulin-Schouleur M, Reperant M, Laurent S, Bree A, Mignon-Grasteau S, Germon P, et al. Extraintestinal pathogenic *Escherichia coli* strains of avian and human origin: Link between phylogenetic relationships and common virulence patterns. J Clin Microbiol. 2007;45:3366–76.
4. Dho-moulin M, Fairbrother JM. Avian pathogenic *Escherichia coli* (APEC). Vet Microbiol. 1999;30:299–316.
5. Dissanayake DRA, Octavia S, Lan R. Population structure and virulence content of avian pathogenic *Escherichia coli* isolated from outbreaks in Sri Lanka. Vet Microbiol. 2014;168:403–12.
6. Adiri RS, Gophna U, Ron EZ. Multilocus sequence typing (MLST) of *Escherichia coli* O78 strains. FEMS Microbiol Lett. 2003;222:199–203.
7. Ewers C, Li G, Wilking H, Kießling S, Alt K, Antáo EM, et al. Avian pathogenic, uropathogenic, and newborn meningitis-causing *Escherichia coli*: How closely related are they? Int J Med Microbiol. 2007;297:163–76.
8. Cloud SS, Rosenberger JK, Fries PA, Wilson RA, Odor EM. In vitro and in vivo characterization of avian *Escherichia coli*. I. Serotypes, metabolic activity, and antibiotic sensitivity. Avian Dis. 1985;29:1084–93.
9. Dozois CM, Fairbrother JM, Harel J. Bossé M. *pap*-and *pil*-related DNA sequences and other virulence determinants associated with *Escherichia coli* isolated from septicemic chickens and turkeys. Infect Immun. 1992;60:2648–56.
10. Mora A, López C, Herrera A, Viso S, Mamani R, Dhabi G, et al. Emerging avian pathogenic *Escherichia coli* strains belonging to clonal groups O111: H4-D-ST2085 and O111: H4-D-ST117 with high virulence-gene content and zoonotic potential. Vet Microbiol. 2012;156:347–52.
11. Olsen RH, Stockholm NM, Permin A, Christensen JP, Christensen H, Bisgaard M. Multi-locus sequence typing and plasmid profile characterization of avian pathogenic *Escherichia coli* associated with increased mortality in free-range layer flocks. Avian Pathol. 2011;40:437–44.
12. Rodriguez-Siek KE, Giddings CW, Doetkott C, Johnson TJ, Fakhr MK, Nolan LK. Comparison of *Escherichia coli* isolates implicated in human urinary tract infection and avian colibacillosis. Microbiology. 2005;151:2097–110.
13. Johnson TJ, Johnson SJ, Nolan LK. Complete DNA sequence of a ColBM plasmid from avian pathogenic *Escherichia coli* suggests that it evolved from closely related ColV virulence plasmids. J Bacteriol. 2006;188:5975–83.
14. Herrero M, De Lorenzo V, Neilands JB. Nucleotide sequence of the *iucD* gene of the pColV-K30 aerobactin operon and topology of its product studied with *phoA* and *lacZ* gene fusions. J Bacteriol. 1988;170:56–64.
15. Parreira VR, Gyles CL. A novel pathogenicity island integrated adjacent to the thrW tRNA gene of avian pathogenic *Escherichia coli* encodes a vacuolating autotransporter toxin. Infect Immun. 2003;71:5087–96.
16. Mo SS, Kristoffersen AB, Sunde M, Nødtvedt A, Norström M. Risk factors for occurrence of cephalosporin-resistant *Escherichia coli* in Norwegian broiler flocks. Prev Vet Med. 2016;130:112–8.
17. Agersø Y, Jensen JD, Hasman H, Pedersen K. Spread of extended spectrum cephalosporinase-producing *Escherichia coli* clones and plasmids from parent animals to broilers and to broiler meat in a production without use of cephalosporins. Foodborne Pathog Dis. 2014;11:740–6.
18. Zurfluh K, Wang J, Klumpp J, Nüesch-Inderbinen M, Fanning S, Stephan R. Vertical transmission of highly similar blaCTX-M-1-harboring IncI1 plasmids in *Escherichia coli* with different MLST types in the poultry production pyramid. Front Microbiol. 2014;5:1–7.
19. Nilsson O, Borjesson S, Landen A, Bengtsson B. Vertical transmission of *Escherichia coli* carrying plasmid-mediated AmpC (pAmpC) through the broiler production pyramid. J Antimicrob Chemother. 2014;69:1497–500.
20. Bisgaard M, Pedersen JR, Christensen H. Danish Commercial Poultry (Dansk Erhvervsfjerkræ). Oct. 2015;10:14–9.
21. Ronco T, Stegger M, Andersen PS, Pedersen K, Li L, Thøfner ICN, et al. Draft genome sequences of two avian pathogenic *Escherichia coli* strains of clinical importance, E44 and E51. Genome Announc. 2016;4:1–2.
22. The National Center For Biotechnology Information. Sequence Read Archive. https://www.ncbi.nlm.nih.gov/sra. Accessed Nov 2016.
23. Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH, et al. Sex and virulence in *Escherichia coli*: an evolutionary perspective. Mol Microbiol. 2006;60:1136–51.
24. Joensen KG, Tetzschner AMM, Iguchi A, Aarestrup FM, Scheutz F. Rapid and easy *in silico* serotyping of *Escherichia coli* isolates by Use of whole-genome sequencing data. Carroll KC, editor. J Clin Microbiol. 2015;53:2410–26.
25. Center For Genomic Epidemiology.http://www.genomicepidemiology.org. Accessed 1 Sept 2016.
26. Northern Arizona SNP Pipeline. http://tgennorth.github.io/NASP/.
27. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26:589–95.
28. Delcher AL, Phillippy A, Carlton J, Salzberg SL. Fast algorithms for large-scale genome alignment and comparison. Nucleic Acids Res. 2002;30:2478–83.
29. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 2011;43:491–8.
30. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 2003;52:696–704.
31. Letunic I, Bork P. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Res. 2011;39:W475–8.
32. The National Center For Biotechnology Information. ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/bacteria/Escherichia_coli/latest_assembly_versions. Accessed 15 Apr 2016.
33. MyDbFinder. https://cge.cbs.dtu.dk/services/MyDbFinder/. Accessed 25 Aug 2016.
34. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, et al. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother. 2012;67:2640–4.
35. Huja S, Oren Y, Trost E, Brzuszkiewicz E, Biran D, Blom J, et al. Genomic avenue to avian colisepticemia. MBio. 2015;6:1–13.
36. Dziva F, Hauser H, Connor TR, van Diemen PM, Prescott G, Langridge GC, et al. Sequencing and functional annotation of avian pathogenic *Escherichia coli* serogroup O78 strains reveal the evolution of *E. coli*

lineages pathogenic for poultry via distinct mechanisms. Infect Immun. 2013;81:838–49.
37. La Ragione RM, Sayers AR, Woodward MJ. The role of fimbriae and flagella in the colonization, invasion and persistence of *Escherichia coli* O78:K80 in the day-old-chick model. Epidemiol Infect. 2000;124:351–63.
38. Gao Q, Wang X, Xu H, Xu Y, Ling J, Zhang D, et al. Roles of iron acquisition systems in virulence of extraintestinal pathogenic *Escherichia coli*: salmochelin and aerobactin contribute more to virulence than heme in a chicken infection model. BMC Microbiol. 2012;12:143.
39. Nolan LK, Horne SM, Giddings CW, Foley SL, Johnson TJ, Lynne AM, et al. Resistance to serum complement, *iss*, and virulence of avian *Escherichia coli*. Vet Res Commun. 2003;27:101–10.
40. Guabiraba R, Schouler C. Avian colibacillosis: Still many black holes. FEMS Microbiol Lett. 2015;362:1–8.

Table S1. APEC- and human UPEC-associated virulence genes

Genes	Description	Accession No	Reference
Adhesins			
fimC	Type 1 fimbriae	CP004009	[1]
papC	Pilus associated with pyelonephritis	AM690766	[2]
fimA	Type 1 fimbriae	NC_000913	[3]
tsh	Temperature-sensitive haemagglutinin	AF218073	[4]
Iron acquisition			
fyuA	Ferric Yersinia uptake	Z38064	[5]
iroN	Catecholate siderophore	DQ381420	[6]
irp2	Iron-repressible protein	L18881	[7]
iucA	Aerobactin	X76100	-
iucD	Aerobactin	M18968	[8]
iroD	Salmochelin siderophore	DQ381420	[6]
Protectins/serum resistance			
cva/cvi	colicin V genes	AJ223631	[9]
iss	Increased serum survival	AF042279	[10]
ompA	Outer membrane protein	CP004009	[1]
Toxins			
vat	Vacuolating autotransporter toxin	AY151282	[11]
hlyA	Haemolysin A	FM180012	[12]
Invasins			
ibeA	Invasion of brain endothelium	CP000468	[13]

Description of APEC- and human UPEC-associated virulence genes [14][15]

References

1. Mangiamele P, Nicholson B, Wannemuehler Y, Seemann T, Logue CM, Li G, et al. Complete genome sequence of the avian pathogenic *Escherichia coli* strain APEC O78. Genome Announc. 2013;1:e0002613.

2. Zdziarski J, Svanborg C, Wullt B, Hacker J, Dobrindt U. Molecular basis of commensalism in the urinary tract: Low virulence or virulence attenuation? Infect. Immun. 2008;76:695–703.

3. Riley M, Abe T, Arnaud MB, Berlyn MKB, Blattner FR, Chaudhuri RR, et al. *Escherichia coli* K-12: a cooperatively developed annotation snapsh–2005. Nucleic Acids Res. 2006;34:1–9.

4. Dozois CM, Dho-Moulin M, Brée A, Fairbrother JM, Desautels C, Curtiss R. Relationship between the Tsh autotransporter and pathogenicity of avian Escherichia coli and localization and analysis of the *tsh* genetic region. Infect. Immun. 2000;68:4145–54.

5. Rakin A, Urbitsch P, Heesemann J. Evidence for two evolutionary lineages of highly pathogenic *Yersinia* species. J. Bacteriol. 1995;177:2292–8.

6. Johnson TJ, Johnson SJ, Nolan LK. Complete DNA sequence of a ColBM plasmid from avian pathogenic *Escherichia coli* suggests that it evolved from closely related ColV virulence plasmids. J. Bacteriol. 2006;188:5975–83.

7. Guilvout I, Mercereau-Puijalon O, Bonnefoy S, Pugsley AP, Carniel E. High-molecular weight protein 2 of *Yersinia enterocolitica* is homologous to AngR of *Vibrio anguillarum* and belongs to a family of proteins involved in non-ribosomal peptide synthesis. J. Bacteriol. 1993;175:5488–504.

8. Herrero M, De Lorenzo V, Neilands JB. Nucleotide sequence of the *iucD* gene of the pCoIV-K30 aerobactin operon and topology of its product studied with *phoA* and *lacZ* gene fusions. J. Bacteriol. 1988;170:56–64.

9. Otto BR, van Dooren SJM, Nuijens JH, Luirink J, Oudega B. Characterization of a hemoglobin protease secreted by the pathogenic *Escherichia coli* strain EB1. J. Exp. Med. 1998;188:1091–103.

10. Chuba PJ, Leon MA, Banerjee A, Palehaudhuri S. Cloning and DNA sequence of plasmid determinant. Mol Gen Genet. 1989;216:287–92.

11. Parreira VR, Gyles CL. A novel pathogenicity island integrated adjacent to the thrW tRNA gene of avian pathogenic *Escherichia coli* encodes a vacuolating autotransporter toxin. Infect. Immun. 2003;71:5087–96.

12. Burgos YK, Pries K, Pestana de Castro AF, Beutin L. Characterization of the alpha-haemolysin determinant from the human enteropathogenic *Escherichia coli* O26 plasmid pEO5. FEMS Microbiol. Lett. 2009;292:194–202.

13. Johnson TJ, Kariyawasam S, Wannemuehler Y, Mangiamele P, Johnson SJ, Doetkott C, et al. The genome sequence of avian pathogenic *Escherichia coli* strain O1:K1:H7 shares strong similarities with human extraintestinal pathogenic *E. coli* Genomes. J. Bacteriol. 2007;189:3228–36.

14. Ewers C, Li G, Wilking H, Kießling S, Alt K, Antáo EM, et al. Avian pathogenic, uropathogenic, and newborn meningitis-causing *Escherichia coli*: How closely related are they? Int. J. Med. Microbiol. 2007;297:163–76.

15. Rodriguez-Siek KE, Giddings CW, Doetkott C, Johnson TJ, Fakhr MK, Nolan LK. Comparison of *Escherichia coli* isolates implicated in human urinary tract infection and avian colibacillosis. Microbiology. 2005;151:2097–110.

Table S2. *E. coli* isolates from chicken farms (see abbreviations below the table)

Name	Farm	House	Sero-type	Virulence genes	Antibiotic resistance genes	MLST	Broiler (B)/ Parrent (P)	State	Source	Coun-try	Sample dates
E1	1	1	O45: H19	*tsh, iucD, ompA, fimA, fimC, iroN,iroD iucA,iss*	-	ST371	B	D	Liver	DK	310315
E2	1	3	O88:H1	*cva, tsh, iucD, iss, iroN,iroD ompA, iucA, fimC,fimA cvi*	*aad, bla, cat, str, sul, tet, dfr*	ST162	B	D	Liver	DK	310315
E3	2	3	O78:H4	*iucD,iss,fimA,ompA,iucA, fimC,iroN,iroD,vat*	-	ST117	B	D	Liver	DK	150415
E4	2	2	O78:H4	*vat, iucD, iss, ompA,iroD iroN, iucA, fimC, fimA*	-	ST117	B	D	Liver	DK	150415
E5	3	3	O88:H1	*cva,cvi, tsh, iucD, iss, iroN, ompA ,iucA, fimC, fimA,iroD*	*aad, bla, cat, str, sul, tet, dfr*	ST162	B	D	Liver	DK	150415
E6	3	3	O78:H4	vat, iucD, iss, iucA,iroD ompA, iroN, fimA, fimC	*bla, sul, tet, dfr*	ST117	B	D	Liver	DK	150415
E7	4	Unkn	O78:H4	*vat, iucD, iss, fimC, iroN, iucA,fimA,iroD, iucA*	-	ST117	B	D	Liver	DK	150415
E8	4	Unkn	O78:H4	*vat, iucD, iss, iroN, iucA, ompA, fimA, fimC,iroD*	*bla*	ST117	B	D	Liver	DK	150415
E9	5	1	O78:H4	*vat, iucD, iss, iroN, iucA, ompA, fimC, fimA,iroD*	-	ST117	P	D	BM	FIN	290115
E10	6	1	O78:H4	*iucD, iss, iroD, fimC, fimA, iucA, ompA, iroN,vat*	-	ST117	P	D	BM	FIN	280115
E11	7	1	O53:H4	*cva, iucD, irp2,iss,vat, iucA, iroN, fyuA, ompA, fimA, fimC,cvi,iroD*	-	ST117	B	D	BM	FIN	130215
E12	8	1	O78:H4	*vat, iucD, iss, iroN, iucA, fimC, fimA, ompA,iroD*	-	ST117	B	D	BM	FIN	060215
E13	9	1	O53:H4	*cva, vat, iucD, irp2,iss, iroN, fimC,, fimA, fyuA, iucA, ompA,cvi,iroD*	*str*	ST117	B	D	BM	FIN	040215
E14	10	1	O53:H4	*cva, iucD, irp2,iss, vat, fimC, fimA, ompA, iroN, iucA, fyuA,cvi,iroD*	*str*	ST117	B	D	BM	FIN	120215
E15	11	1	O53:H4	*cva, iucD, irp2, iss,iroD vat, fimA, iroN iucA,cvi , fyuA, ompA, fimC*	*str*	ST117	B	D	BM	FIN	120215
E16	12	1	O78:H4	*vat, iucD, iss, iucA, iroN, fimC, fimA, ompA,iroD*	-	ST117	B	D	BM	FIN	130215
E17	13	1	O53:H4	*cva, iucD, irp2,iss, vat, fimC iroN,fyuA,cvi,iroD ompA,iucA,fimA*	*str*	ST117	B	D	BM	FIN	130215
E18	14	1	O78:H4	*vat, iucD, iss, iroN,iroD ompA, fimC, fimA, iucA*	-	ST117	B	D	BM	FIN	130215
E19	15	1	O50O2: H6	*cva, vat, tsh, iucD, irp2, iss, iroN, ompA, fimA,cvi fimC, ibeA, fyuA, iucA,iroD*	*bla*	ST141	B	D	Liver	DK	130515
E20	16	2A	O78:H4	*vat, iucD, iss, fimC, iroN, fimA, ompA, iucA,iroD*	-	ST117	B	D	Liver	DK	130515

Table S2.

Name	Farm	House	Sero-type	Virulence genes	Antibiotic resistance genes	MLST	Broiler (B)/ Parrent (P)	State	Source	Coun-try	Sample dates
E21	16	1B	O78:H4	*irp2,iss, ompA, fimC, fimA, fyuA, iroN,iroD,iucD*	*tet*	ST23	B	D	Liver	DK	130515
E22	17	1	O78:H4	*iss, vat, iucD,, fimA, fimC, iucA, iroN, ompA,iroD*	-	ST117	B	D	Liver	DK	230615
E23	18	1	O78:H4	*iss, iucD, vat, fimC, fimA, ompA, iroN, iucA,iroD*	-	ST117	B	D	Liver	DK	150715
E24	19	1	O161: H4	*iss, irp2, vat, iucD, iroN, ompA, fyuA, iucA, fimA, fimC,papC,iroD*	*bla, tet*	ST117	B	D	Liver	DK	180815
E25	20	1	O18ac: H7	*irp2, tsh,vat,iroD iucD,cva,cvi,ibeA,iss fyuA, fimC, fimA, iucA, iroN, ompA,*	-	ST95	P	D	Unkn	NO	230915
E26	21	Unkn	O78:H4	*iss, vat,iucD, iucA, iroN, fimC, fimA, ompA,iroD*	-	ST117	P	D	Femur	NO	260815
E27	22	1	O78:H4	*iss,vat, iucD, iroN, fimC, fimA, ompA, iucA,iroD*	*bla*	ST117	B	D	Spleen	NO	070915
E28	21	Unkn	O18ac: H7	*iucD, irp2, cvi, cva, ibeA, fyuA, iroD, iucA, fimC, fimA, ompA,tsh,vat, iroN*	*bla*	ST95	P	D	Spleen	NO	130915
E29	23	1	O78:H4	*iss, iucD, vat,, iucA, iroN, ompA, fimC, fimA,iroD*	-	ST117	B	D	Liver	NO	160915
E30	24	1	H23	*iss,cva, cvi, iroN, fimC, ompA, fimA,iroD*	-	ST1163	B	D	Unkn	NO	040815
E31	25	1	O78:H4	*iucD, vat, iroN, fimC,iss fimA, iucA, ompA,iroD*	*bla*	ST117	B	D	Unkn	NO	060815
E32	26	1	O78:H4	*iss, iucD, vat ,fimC, fimA, iroN, ompA, iucA,iroD*	*bla*	ST117	B	D	Unkn	NO	170815
E33	27	1	O1:H7	*iroN, iucA, ompA,cva,cvi fyuA,irp2 fimA, fimC, iroD,papC,iucD,vat,iss*	-	ST95	P	D	Unkn	NO	100815
E34	28	Unkn	O149: H23	*cva,, cvi, iss, iroN, fimC, ompA,iroD,fimA,iss*	-	ST1163	Unkn	D	Femur	NO	160915
E35	29	1	O78:H4	*iss, iucD, vat, iroN, fimC, fimA, iucA, ompA,iroD*	-	ST117	B	D	Liver	DK	230915
E36	29	2	O78:H4	*iss, iucD ,vat,iroN, fimC, fimA ompA, iucA,iroD*	-	ST117	B	D	Liver	DK	230915
E37	30	1	O1:H7	*iss, irp2, iucD, cvi, cva, vat, ompA,iroN, fyuA,iroD iucA, fimC, fimA,papC*	-	ST95	B	D	Liver	DK	061015
E38	30	2	O174: H42	*irp2, iucD, cva, tsh, cvi, iroN, fimA, fimC, iucA, ompA,fyuA,iroD,iss*	*tet*	ST648	B	D	Liver	DK	061015
E39	31	1	O78:H4	*iss, iucD, iroN, iucA,fimC, fimA, ompA,vat,iroD*	-	ST117	P	D	BM	FIN	070815
E40	32	1	O78:H4	*iss,iucD, vat,,iroN, iucA, ompA, fimC, fimA,iroD*	-	ST117	P	D	BM	FIN	030815

Table S2.

Name	Farm	House	Sero-type	Virulence genes	Antibiotic resistance genes	MLST	Broiler (B)/ Parrent (P)	State	Source	Coun-try	Sample dates
E41	33	1	O18ac: H7	*iucD, cva, cvi, fimC, tsh, fimA, iucA iroN, ibeA,iss irp2,iroD ompA,vat,*	-	ST95	P	D	BM	FIN	240615
E42	34	1	O18ac: H7	*irp2, cvi, cva, iucD, fimC, fimA, iroN, ompA, fyuA, tsh iucA, ibeA,vat,iroD,iss*	-	ST95	P	D	BM	FIN	280415
E43	35	1	O18:H7	*iucD, irp2, cva, cvi, fimC, fimA, fyuA, ompA, ibeA, iucA,tsh,iroN,iroD,iss*	-	ST95	P	D	BM	FIN	260515
E44	36	1a	O78:H4	*iss, iucD, vat, iroN, fimC, fimA, ompA, iucA,iroD,*	-	ST117	P	D	Liver	DK	200415
E45	36	1b	O78:H4	*iss, iucD, vat, iroN,iroD ompA, iucA, fimA, fimC*	-	ST117	P	D	Liver	DK	200415
E46A	36	2a	H4	*iucD, iss, iroN, fimC, ompA, fimA, iroD, iucA,vat*	-	ST117	P	D	Liver	DK	200415
E47	36	2b	O78:H4	*iss, vat, iucD, iroD, iroN, ompA, iucA, fimA, fimC*	-	ST117	P	D	Liver	DK	200415
E48	36	P1	O78:H4	*iss, iucD, vat,iroD iroN, iucA, ompA, fimA, fimC*	-	ST117	P	D	Liver	DK	200415
E49	36	P2	O78:H4	*iss, iucD, vat, iroN, iucA, ompA, fimC, fimA,iroD*	-	ST117	P	D	Liver	DK	200415
E50	36	P3	O78:H4	*iss iucD, vat, ompA,iroD iroN, fimA, iucA, fimC,*	-	ST117	P	D	Liver	DK	200415
E51	37	3a	O2:H5	*cva, cvi, irp2, iucD, tsh, iroN, ompA, ibeA, fimC, fyuA, fimA, iucA,iroD,iss*	-	ST140	P	D	Liver	DK	210115
E52A	37	3b	O78:H4	*vat, iucD, fimC, fimA,iss ompA, iucA, iroN,iroD*	-	ST117	P	D	Liver	DK	210115
E53	38	1	H4	*iss, iucD, vat, iroN, fimC, ompA, fimA, iucA,iroD*	-	ST117	B	D	Liver	DK	2015
E54A	38	2	H4	*vat, iucD, iss, fimC, fimA, ompA, iucA, iroN,iroD*	-	ST117	B	D	Liver	DK	2015
E56A	40	1	O78:H4	*vat, iucD, iroN, iroD ompA,iss iucA, fimA, fimC,*	-	ST117	P	D	Liver	DK	170415
E57	40	2	O78:H4	*iss, vat, iucD, fimC, iucA, fimA, ompA, iroN,iroD*	*aad, bla, sul, dfr*	ST117	P	D	Liver	DK,	170415
E58	41	1	O78:H4	*tsh, irp2, cvi, cva,iss iucD, fimA, iucA,fyuA, fimC iroN,iroD, ompA,*	*Aad, sul, tet*	ST23	B	D	Liver	DK	300415
E59	42	1	O78:H4	*iss, iucD, vat, iroN, fimC, fimA, ompA, iucA,iroD*	-	ST117	B	D	Liver	DK	100515

Table S2.

Name	Farm	House	Sero-type	Virulence genes	Antibiotic resistance genes	MLST	Broiler (B)/ Parrent (P)	State	Source	Coun-try	Sample dates
E60	43	1	O78:H4	*iss, iucD, vat, iroN,iroD ompA, iucA, fimA, fimC*	-	ST117	B	D	Liver	DK	280415
E61	44	2	O50/O2 :H5	*cva,cvi,irp2.tsh,papC iroN,iucA,fyuA,ibeA, fimC,fimA,ompA,vat, iss,iroD*	-	ST95	B	D	Liver	DK	090515
E62	44	1	O50/O2 :H5	*cva, vat, iucD, irp2, cvi,, ibeA, fimC, fimA, iucA, ompA, fyuA, iroN,iroD*	-	ST355	B	D	Liver	DK	090515
E63	45	1	O78:H4	*iss, vat iucD, fimA, iucA, iroN, ompA, fimC,iroD*	-	ST117	B	D	Liver	DK	280415
E64	46	2	H10	*cva, iss, iucD, iroN,iroD, ompA,iucA tsh, cvi*	-	ST93	B	D	Liver	DK	210415
E65	47	1	O78:H4	*iss, iucD, vat, iucA, iroN, fimC, fimA, ompA,iroD*	-	ST117	B	D	Liver	DK	120515
E66A	47	2	O78:H4	*vat, iucD, fimC, fimA,,iss iroN, ompA, iucA,iroD*	-	ST117	B	D	Liver	DK	120515
E67A	48	1	H4	*iss, irp2, iucD, vat,, iroN, fyuA, fimC, fimA, iucA, ompA,iroD*	*aad, bla, str, sul, tet, dfr*	ST117	B	D	Liver	DK	280415
E68A	48	2	H4	*iss, irp2, iucD, vat,, iroN, ompA, fyuA, iucA, fimA, fimC,iroD*	*aad, bla, str, sul, tet, dfr*	ST117	B	D	Liver	DK	280415
E69	49	Unkn	O78:H4	*iucD,iss,fimC,iroN, iroD,vat, iucA, ompA,fimA*	-	ST117	Unkn	D	PT	DK	171115
E72	50	Unkn	H4	*cva,cvi,fimA,fimC,iucA iroD,irp2,iss,iucD, iroN, ompA,vat,iss*	-	ST117	Unkn	D	PC	DK	171115
E75	51	Unkn	O53:H4	*cva,cvi,iroD,iss,irp2, iucD,iroN,fimC,fimA, ompA,fyuA,iucA,vat*	*str*	ST117	Unkn	D	Liver	DK	231115
E77	52	Unkn	O78:H4	*iucD,iss,iucA,fimA, iroD,vat,iroN,ompA, fimC*	-	ST117	Unkn	D	Femur	DK	301115
E79	53	Unkn	O5:H10	*cva,cvi,ompA,fimA, iroD,iss,iucD,iroN, iucA,tsh,ibeA*	-	ST93	Unkn	D	PT	DK	281015
E80	54	Unkn	O78:H4	*iucD,iss,ompA,iroN, iucA,vat,fimA,fimC, iroD*	-	ST117	Unkn	D	Liver	DK	261115
E86	55	Unkn	O78:H4	*iucD,iss,vat,iroN,iroD, ompA,fimC,fimA,iucA*	-	ST117	P	D	Liver	DK	090116
E87	56	1	O53: H18	*cvi,iss,cva, iss,ompA,iroN,fimC, iroD*	*aad bla, sul, tet, dfr*	ST1638	B	D	Liver	DK	070116
E88	57	Unkn	H9	*fimA,fimC,ompA*	-	ST38	B	D	Liver	DK	201215
E89	43	Unkn	H10	*cva,cvi,irp2,fyuA, ompA,fimC,fimA,iss*	*aad*	ST10	B	D	Liver	DK	111215

Table S2.

Name	Farm	House	Sero-type	Virulence genes	Antibiotic resistance genes	MLST	Broiler (B)/ Parrent (P)	State	Source	Coun-try	Sample dates
E90	58	Unkn	O53:H4	*cva,cvi,iss,irp2,iucD, iroN,ompA,fyuA,iucA, fimA,fimC,vat,iroD*	*str*	ST117	B	D	Liver	DK	110116
E91	59	Unkn	O78:H4	*iucD,iss,iroD,fimA iucA,vat,iroN,fimC, ompA*	-	ST117	B	D	Liver	DK	041215
E92	60	Unkn	O5:H10	*cva,cvi, iss,iucD,iroD, iroN,iucD,ompA,tsh*	-	ST93	B	D	Liver	DK	071215
E93	57	Unkn	O18:H7	*cva,cvi,iroD,iss, iroN,fimA,fimC,ompA*	-	Unkn	B	D	Liver	DK	111215
E94	61	Unkn	O78:H4	*iucD,iss,iroD vat,iroN,ompA,iucA, fimA,fimC*	-	ST117	B	D	Liver	DK	151215
E95	62	Unkn	O78: H49	*fimA,fimC,ompA,iss*	-	ST2248	Unkn	D	Liver	DK	1. quarter 2016
E96	63	Unkn	O113: H48	*fimC,ompA,iss*	*str, aad, bla, sul, tet, dfr*	ST10	Unkn	D	Liver	DK	1. quarter 2016
E97	28	Unkn	O149: H23	*cva,cvi,iroN,ompA, fimC,fimA,iroD,iss*	-	ST1163	B	D	Femur	NO	160915
E98	64	Unkn	O78:H4	*iucD,iss,iroD vat,iroN,ompA,iucA, fimA,fimC*	-	ST117	B	D	Femur	NO	161015
E99	65	Unkn	O50/O2 :H5	*cva,cvi,iroD,iss irp2,iucD,tsh,iroN,ibe A,fyuA,ompA,iucA,fim A,fimC*	-	ST140	B	D	PC	NO	231115
E100	66	Unkn	O1:H42	*cva,cvi, ,iucA tsh,iss irp2,iucD,fimC, fimA, ,ompA,fyuA,iroD,iroN*	*bla*	ST648	P	D	Unkn	NO	261015
E101	67	Unkn	O103: H21	*cva, cvi,iroD,iss, iucD,iroN,iucA,ompA,f imc,fimA,tsh*	-	ST101	P	D	Liver	NO	250115
E102	68	Unkn	O78:H4	*iucD,iss,iroD, vat,fimC,fimA,iroN, ompA,iucA*	-	ST117	B	D	Unkn	NO	270915
E104	69	Unkn	O75: H42	*ompA,fimC,fimA*	*str*	ST2223	B	D	BM	DK	2015
E105	70	Unkn	O78:H4	*iucD,iss,iroD vat,iucA,iroN,fimC, fimA,ompA*	-	ST117	B	D	BM	DK	2015
E106	71	Unkn	O18: H49	*cva,cvi,iroD iss,iroN,ompA,fimC, fimA*	-	ST212	B	D	BM	DK	2015
E107	72	3	O53: H4	*cva,cvi,iroD iss,irp2,iucD,iroN, ompA,,fyuA,iucA,fimA, fimC,vat*	*dfr, sul, str, aad*	ST117	B	D	BM	DK	2015
E109	Unkn	3	O78:H4	*iucD,iss,iroD vat,ompA,iroN,iucA, fimA,fimC*	-	ST117	B	D	BM	DK	2015

Table S2.

Name	Farm	House	Sero-type	Virulence genes	Antibiotic resistance genes	MLST	Broiler (B)/ Parrent (P)	State	Source	Coun-try	Sample dates
E110	43	2	O78:H4	*iucD,iss,iroD vat,iroN,fimC,fimA,iuc A,ompA*	-	ST117	B	D	BM	DK	2015
E111	73	2	O149: H23	*cva,,cvi,iroN,fimC,omp A,iroD,iss*	*aad, bla, sul, dfr*	ST1163	B	D	BM	DK	2015
E112	74	3	O8:H8	*cva,cvi,iss,fimA,fimC ompA,iroD,iroN*	-	ST109	P	H	Cloaca	DK	July 2015
E113	75	1	O103: H2	*cva,cvi,iucD,ompA,iuc A,fimC,fimA,iss*	*sul, bla*	ST1146	P	H	Cloaca	DK	July 2015
E114	76	1	O76: H19	*fimA,ompA,fimC*	-	ST675	P	H	Cloaca	DK	July 2015
E116	77	1	O103: H2	*cva,cvi,iucD,iucA,fimA fimC,ompA*	*sul, bla*	ST1146	P	H	Cloaca	DK	July 2015
E117	78	1	O82: H10	*ompA,fimA,fimC,iss*	-	ST5625	P	H	Cloaca	DK	July 2015
E118	79	1	O149: H10	*cva,cvi,iss,iroN,iroD,o mpA,fimC,,iss*	-	ST746	P	H	Cloaca	DK	July 2015
E119	80	1	O28ac/ O42: H21	*cva,cvi,fimA,ompA, fimC*	-	ST3714	P	H	Cloaca	DK	July 2015
E120	81	Unkn	O69: H11	*ompA*	-	ST10	P	H	Cloaca	DK	July 2015
E121	39	Unkn	O103: H2	*Cva,cvi,iucD, fimA,fimC,iucA,ompA*	*sul, bla*	ST1146	P	H	Cloaca	DK	July 2015
E122	82	Unkn	O103: H2	*cva,cvi,iucD,fimA,fim C,iucA,ompA*	*sul, bla*	ST1146	P	H	Cloaca	DK	July 2015
E123	83	Unkn	O9:H21	*cva,cvi,fimC,fimA, ompA*	-	ST1642	P	H	Cloaca	DK	July 2015
E124	84	Unkn	O8:H20	*cva,cvi, iucD,iucA,ompA*	-	ST155	B	H	Cloaca	DK	100815
E125	85	Unkn	O17/O7 7:H18	*iroN,ompA,fimA,fimC ,iucD,cva,cvi,iss*	-	ST69	B	H	Cloaca	DK	100815
E126	86	Unkn	O15: H18	*cva,cvi,irp2,iss,iucD, fimC,fimA,iucA,ompA, fyuA,iroN,iroD*	-	ST69	B	H	Cloaca	DK	170815
E127	87	Unkn	O103:H 2	*cva,cvi,iucD,ompA,fim A,fimC,iucA,iss*	*sul, bla*	ST1146	B	H	Cloaca	DK	100815
E128	Unkn	Unkn	O50/O2 :H5	*cva,cvi,iss,iucD,irp2,fy uA,ompA,iroN,ibeA,iu cA,fimC,fimA,tsh,iroD*	-	ST140	B	D	Liver	PO	1. quarter 2016
E129	Unkn	Unkn	O120: H4	*cva,cvi,iss,vat,irp2,om pA,fyuA,iroN,ibeA, iroD,fimA,fimC,iucD*	*sul*	ST428	B	D	Liver	PO	1. quarter 2016

Abbreviations

B: Broiler chickens,

P: Parents (layer hens)

D: Diseased

H: Healthy

BM: Bone Marrow

PT: Peritoneum

PC: Pericardia

Unkn: Unknown

Table S3. ST117 *E. coli* strains from NCBI

Strain	Host	Collection date	Country	Isolation source	GenBank Acc. no.
H299	-	-	-	-	NZ_GL884490.1
SEPT362	Chicken	Mar-1994	Brazil	Liver	NZ_AOGL00000000
HVH 79	Human	2003	Denmark	Blood	NZ_KE699539.1
HVH 188	Human	2003	Denmark	Blood	NZ_KE700731.1
897	-	-	-	-	NZ_AYQF00000000
53C	Chicken	May-2010	Netherlands	Retail meat	NZ_AYRA00000000
53C.1	Chicken	May-2010	Netherlands	Retail meat	NZ_JXMX00000000
1047	-	-	-	-	NZ_AYQG00000000
38.52	-	-	-	-	NZ_AYQH00000000
03-3458	-	-	USA	-	NZ_JHNV00000000
2-177-06_S4_C1	Human	Jul-2009	Tanzania	Stool	NZ_JNQG00000000
Blood-10-0682	Human	2010	USA	Blood	NZ_JSQL00000000
EC5	Chicken	Sep-2013	Malaysia	Lever	NZ_JWKF00000000
EC7	Chicken	Sep-2013	Malaysia	Spleen	NZ_JWKG00000000
BIDMC101	Human	2014	-	-	NZ_KQ087895.1
2009C-3133	Human	-	-	Stool	NZ_CP013025
GN02004.	Human	-	USA	Body fluid	NZ_LQRU00000000
Cattle5	Cattle	1980	France	Feces	NZ_LVLO00000000
Cattle12	Cattle	2011	China	Feces	NZ_LVLV00000000
Cattle14	Cattle	2011	China	Feces	NZ_LVLX00000000
Cattle19	Cattle	2011	China	Feces	NZ_LVMC00000000

Table S4. Assembly metrics of *E. coli* isolates

Isolate	Average coverage (×)	N50 (Kb)	Number of contigs	Assembly size (Mb)
E1	48	95.5	214	5.3
E2	100	184.6	151	5.0
E3	43	59.9	284	5.1
E4	49	99.8	178	5.1
E5	18	32.6	356	5.0
E6	47	112.0	213	5.2
E7	41	74.7	226	5.1
E8	65	108.3	181	5.1
E9	33	67.9	251	5.2
E10	34	77.4	372	5.1
E11	35	75.7	188	5.1
E12	37	52.3	290	5.2
E13	33	46.1	467	5.3
E14	30	64.0	256	5.1
E15	54	123..0	150	5.1
E16	45	87.5	245	5.2
E17	52	99.4	152	5.1
E18	38	58.3	297	5.1
E19	39	128.0	144	5.2
E20	44	65.9	261	5.1
E21	52	128.3	132	5.0
E22	48	112.1	238	5.3
E23	46	115.1	174	5.1
E24	53	166.2	181	5.2
E25	36	107.2	162	5.2
E26	53	97.9	182	5.1
E27	63	87.3	209	5.1
E28	55	162.5	137	5.1
E29	66	87.4	233	5.2
E30	58	85.5	193	5.0
E31	59	72.1	283	5.2
E32	65	93.5	193	5.2
E33	56	85.8	182	5.0
E34	72	110.7	132	5.1
E35	114	87.4	235	5.3
E36	68	87.8	223	5.2
E37	99	188.6	142	5.1
E38	51	90.3	235	5.3
E30	63	72.2	239	5.2
E40	94	97.5	221	5.1

Table S4.

Isolate	Average coverage (×)	N50 (Kb)	Number of contigs	Assembly size (Mb)
E41	91	135.4	137	5.1
E42	46	56.3	229	5.1
E43	75	155.0	119	5.2
E44	91	83.8	195	5.1
E45	86	120.1	204	5.1
E46A	>15	-	354	5.2
E47	83	111.4	206	5.2
E48	92	119.6	179	5.2
E49	70	99.3	185	5.2
E50	61	98.8	207	5.1
E51	58	100.1	217	5.2
E52A	>15	-	366	5.1
E53	53	64.0	270	5.1
E54A	>15	-	344	5.1
E56A	>15	-	306	5.2
E57	55	85.9	252	5.4
E58	69	85.0	192	5.1
E59	24	35.3	373	5.2
E60	57	97.4	193	5.1
E61	39	82.7	262	5.3
E62	47	83.7	231	5.1
E63	85	92.4	204	5.3
E64	36	42.8	357	4.7
E65	125	130.6	160	5.1
E66A	>15	-	353	5.1
E67A	>13	-	288	5.1
E68A	>15	-	244	5.1
E69	27	55.6	345	5.1
E72	77	112.0	256	5.2
E75	33	75.8	196	5.1
E77	67	96.8	184	5.1
E79	42	38.2	891	5.5
E80	34	73.9	201	5.1
E86	29	50.3	333	5.2
E87	44	58.9	207	4.8
E88	37	91.0	175	4.9
E89	37	63.5	221	4.9
E90	36	98.3	176	5.1
E91	28	62.8	258	5.2

Table S4

Isolate	Average coverage (×)	N50 (Kb)	Number of contigs	Assembly size (Mb)
E92	33	58.9	188	4.9
E93	25	41.6	580	5.3
E94	49	119.6	159	5.2
E95	33	84.1	207	5.3
E96	119	68.3	229	5.1
E97	43	123.3	150	5.1
E98	113	120.2	173	5.1
E99	110	107.6	177	5.3
E100	128	129.9	185	5.5
E101	121	110.3	148	5.0
E102	105	111.4	186	5.1
E104	88	84.7	163	5.1
E105	43	95.7	189	5.1
E106	114	116.3	151	5.0
E107	74	97.6	186	5.2
E109	109	108.9	177	5.1
E110	95	112.3	226	5.1
E111	114	109.2	185	5.1
E112	85	96.6	143	5.0
E113	36	76.5	185	5.1
E114	116	102.3	159	5.1
E116	141	103.8	158	5.1
E117	72	75.0	146	4.7
E118	117	112.2	178	4.9
E119	111	94.0	121	4.9
E120	111	53.2	241	4.7
E121	73	92.2	168	5.0
E122	72	94.6	152	5.1
E123	116	90.0	178	5.0
E124	80	61.9	264	5.3
E125	100	128.1	163	5.2
E126	138	140.8	119	5.2
E127	76	97.2	161	5.1
E128	113	106.4	188	5.2
E129	78	137.6	121	5.2

Appendix III

Paper III

Ronco T, Stegger M, Ng KL, Lilje B, Lyhs U, Andersen PS, Pedersen K. 2017. Genome analysis of *Clostridium perfringens* isolates from healthy and necrotic enteritis infected chickens and turkeys. BMC Res. Notes 10, 1–6.

Ronco *et al. BMC Res Notes (2017) 10:270*
DOI 10.1186/s13104-017-2594-9

BMC Research Notes

RESEARCH NOTE

Open Access

Genome analysis of *Clostridium perfringens* isolates from healthy and necrotic enteritis infected chickens and turkeys

Troels Ronco[1*], Marc Stegger[2], Kim Lee Ng[2], Berit Lilje[2], Ulrike Lyhs[1], Paal Skytt Andersen[2,3] and Karl Pedersen[1]

Abstract

Objective: *Clostridium perfringens* causes gastrointestinal diseases in both humans and domestic animals. Type A strains expressing the NetB toxin are the main cause of necrotic enteritis (NE) in chickens, which has remarkable impact on animal welfare and production economy in the international poultry industry. Three pathogenicity loci NELoc-1, -2 and -3 and a collagen adhesion gene *cnaA* have been found to be associated with NE in chickens, whereas the presence of these has not been investigated in diseased turkeys. The purpose was to investigate the virulence associated genome content and the genetic relationship among 30 *C. perfringens* isolates from both healthy and NE infected chickens and turkeys, applying whole-genome sequencing.

Results: NELoc-1, -3, *netB* and *cnaA* were significantly associated with NE isolates from chickens, whereas only NELoc-2 was commonly observed in both diseased turkeys and chickens. A putative collagen adhesion gene that encodes a von Willebrand Factor (vWF) domain was identified in all diseased turkeys and designated as *cnaD*. The phylogenetic analysis based on single nucleotide polymorphisms showed that the isolates generally were not closely related. These results indicate that virulence factors and pathogenicity loci associated with NE in chickens are not important to the same extent in diseased turkeys except for NELoc-2. A putative collagen adhesion gene which potentially could be of importance in regard to the NE pathogenesis in turkeys was identified and need to be further investigated. Thus, the pathogenesis of NE in turkeys appears to be different from that of broiler chickens.

Keywords: Genome analysis, Necrotic enteritis, Poultry, Virulence factors

Introduction

Clostridium perfringens is a Gram-positive anaerobic bacterium that causes gastrointestinal diseases in humans and domestic animals [1]. Virulence is primarily due to production of various types of extracellular toxins, and *C. perfringens* strains are assigned a specific toxin type (A–E) dependent on which major toxins (α, β, ε, ι) they produce [2, 3]. Colonization of the intestine by *netB*-positive type A strains is known to be the main cause of necrotic enteritis (NE) in broilers [4, 5], which constitutes a considerable burden to the animal welfare and production yield in the international poultry industry [6]. NetB is essential in the pathogenesis of NE in chickens [4, 5], whereas the prevalence of *netB* is low among diseased turkeys [7, 8]. Few studies of the virulence gene content in strains from turkeys with NE and enteric diseases have been carried out [9] and to our knowledge, there are no reports on whole-genome sequencing of *C. perfringens* isolates from turkeys.

netB is found on a 42 kb plasmid-encoded and NE associated pathogenicity locus called NE locus-1 (NELoc-1), which also harbors several other virulence genes [10]. Two other loci have also been found to be associated with NE in chickens, NELoc-2 (11.2 kb) and NELoc-3 (5.6 kb), which are chromosomally and plasmid-encoded respectively [10]. Furthermore, a recent study showed that a

*Correspondence: troro@vet.dtu.dk
[1] National Veterinary Institute, Technical University of Denmark, Bülowsvej 27, 1870 Frederiksberg C, Denmark
Full list of author information is available at the end of the article

collagen adhesion gene, *cnaA* is involved in NE in chickens [11].

The purpose of this study was to investigate the virulence associated genome content and genetic relationship among *C. perfringens* isolates from healthy and NE afflicted chickens and turkeys, applying whole-genome sequencing.

Main text

Clostridium perfringens isolates

Isolates were sampled from healthy (n = 4) and diseased (n = 13) chickens and from healthy (n = 4) and diseased (n = 9) turkeys. Both chickens and turkeys were from conventionally raised indoor flocks. The samples were collected on 14 different Danish chicken farms between 1997 and 2002 [12], and seven different Finnish turkey farms between 1997 and 2010 [8]. The samples were primarily from the intestine, but six chicken isolates were obtained from liver samples (Additional file 1).

DNA purification and sequencing

Colonies were grown overnight on Columbia agar base (Oxoid, CM0331, Hampshire, UK) supplemented with 5% calf blood (Statens Serum Institut, Copenhagen, Denmark) at 37 °C under anaerobic conditions (10% CO_2, 10% H_2 and 80% N_2). Single colonies were cultured in 5 ml Trypticase Soy Broth (Becton–Dickinson, Franklin Lakes, N. Jersey,) under same anaerobic conditions, and DNA was purified using the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions. Subsequently, DNA libraries were built using the Nextera XT kit (Illumina Inc., San Diego, Ca) according to manufacturer's instructions, and subjected to whole-genome sequencing using Illumina's MiSeq platform with paired-end read lengths of 2 × 251 bp.

Assembly and identification of preselected *C. perfringens* genes

All reads were deposited in the NCBI SRA [13] (Additional file 1) and *de novo* assembled using CLCbio's Genomics Workbench (GW) v6.5 (CLCbio's, Aarhus, Denmark) on default settings and a minimum contig-size of 500 NTs. Subsequently, various types of genes were identified using the web-tool MyDbFinder v1.1 [14] with selected verification of open reading frames (ORFs) using the BLASTN implementation in CLCbio's Main Workbench (MW) v8.5. Nucleotide sequences of various genes were obtained from the Virulence factor database [15] and the NCBI nucleotide database (Additional file 2). All isolates were multi-locus sequence typed (MLST) at PubMLST [16].

SNP calling and investigation of the general gene content

Single nucleotide polymorphisms (SNPs) were identified using CSI Phylogeny v1.4 [17] on default settings and with *C perfringens* strain ATCC 13124 (NCBI Accession No. NC_008261) as reference chromosome. Thus, all SNPs had a minimum depth of ≥10×, a quality of ≥30 and a distance of ≥10 to the next SNP. The phylogenetic tree was further modified in iTOL v3.1 [18].

Investigation of the general gene content was performed on the assembled contigs using Prokka v1.10 [19] (default settings) and Roary v3.2.4 [20] (–e–mafft settings) for gene detection and annotation, followed by core and accessory genome identification, respectively.

Statistics

Statistical analyses were carried out using GraphPad Prism v5.02 (GraphPad Software Inc., San Diego, Ca). Differences in the presence of virulence genes and loci between isolates from diseased chickens and turkeys were investigated using a Fisher's exact test and considered significant when $p < 0.05$.

Results

Of the 30 isolates, 21 were sequenced to an average coverage of >80× whereas the rest had an average coverage of >40× except isolate T43 that had >30×. The ORFs were generally determined with thresholds of 90% nucleotide identity and 90% coverage of query sequence length. All isolates were confirmed to be of toxin type A as they only carried the *plc* gene (encoding α-toxin) (Table 1) and MLST showed that all isolates were of unknown sequence type (STs), except two isolates (C26 and C31) from diseased chickens that were of ST21.

netB was only found in isolates from diseased poultry and primarily in chickens. Of the NE isolates from chickens, 77% (10/13) were *netB*-positive, whereas only 22% (2/9) of the NE isolates from turkeys carried *netB* ($p = 0.0274$) (Table 1). The NELoc-1 and -3 associated genes (Additional file 3), were primarily observed among isolates from diseased chickens (Table 1). The isolates could be divided into two types of groups. A high prevalence group (HPG) that on average carried 94% of the NELoc-1 genes, and a low prevalence group (LPG) that on average carried 10% of the genes (Table 1). As with NELoc-1, the isolates were divided into a HPG that on average carried 96% of the NELoc-3 genes, and a LPG that on average carried 24% of the NELoc-3 genes (Table 1). The NELoc-1 HPG included 77% (10/13) of NE isolates from chickens and 22% (2/9) of the NE isolates from turkeys ($p = 0.0274$). The NELoc-3 HPGs had 77% (10/13) of NE isolates from chickens and in a single isolate (1/9) from a diseased turkeys ($p = 0.0075$).

Table 1 Virulence genes and loci identified in 30 *C. perfringens* isolates

Isolate	State	Type	*cnaD*	*cnaA*	*netB*	NELoc-1	NELoc-2	NELoc-3
C1	H	A	+			0	0	0
C3	H	A				12	18	20
C7	H	A	+			3	*100*	20
C8	H	A				9	*100*	20
C24	D	A		+	+	*97*	*100*	*100*
C25	D	A		+		21	*100*	*100*
C26	D	A		+	+	*100*	*100*	*100*
C27	D	A	+			0	0	0
C31	D	A		+	+	*100*	*100*	*100*
C32	D	A	+			18	0	20
C33	D	A	+	+	+	*88*	*100*	*80*
C36	D	A		+	+	*100*	*100*	*100*
C37	D	A			+	*100*	*100*	60
C41	D	A			+	*100*	*100*	*100*
C48	D	A	+	+	+	*94*	*100*	*100*
C124	D	A	+	+	+	*100*	*100*	*80*
C125	D	A			+	*100*	*100*	*100*
T1	D	A	+			0	*100*	0
T5	D	A	+			27	*100*	20
T6	D	A	+			27	0	40
T11	D	A	+	+	+	*100*	*100*	*100*
T14	D	A	+			27	*100*	40
T16	D	A	+			9	0	20
T46	D	A	+			12	*100*	40
T53	D	A	+		+	*100*	0	40
T84	D	A	+			12	*100*	40
T18	H	A	+			21	0	60
T22	H	A		+		9	0	40
T34	H	A				9	0	20
T43	H	A	+			12	*100*	20

The + mark indicates gene presence in isolates from healthy (H) or diseased (D) chickens (C) and turkeys (T). The prevalence (in %) of NELoc-1, -2 and -3 genes are shown. High prevalence of genes is in italics

In contrast to the NELoc-1 and -3 genes, the NELoc-2 genes (Additional file 4), when identified in an isolate, were all conserved except in isolate C3 that carried 18% of the genes (Table 1). NELoc-2 was found in 85% (11/13) of the NE isolates from chickens, and in 67% (6/9) of the NE isolates from turkeys ($p = 0.6090$). Only two isolates from healthy chickens and a single isolate from a healthy turkey carried NELoc-2 (Table 1). Sixty-two % (8/13) of the isolates from diseased chickens carried *cnaA* whereas this gene was found only in one diseased turkey ($p = 0.0306$). A single healthy turkey carried *cnaA* (Table 1).

The SNP analysis included 50,643 variant positions and 61.9% of the reference genome was covered by all isolates. A phylogenetic analysis revealed no specific clustering among the poultry isolates which in overall were not closely related (Fig. 1). The investigation of the general gene content showed a putative collagen adhesion gene of 2787 nucleotides (NTs) (Additional file 5) here labelled *cnaD*, which was present in all isolates from diseased turkeys (9/9) but only in 39% (5/13) of diseased chickens ($p = 0.0055$). In healthy birds, *cnaD* was found in half (2/4) of both the turkeys and chickens (Table 1). The *cnaD* gene encodes a 928 long amino acid (AA) sequence (Additional file 5), which was analyzed using BLASTP v2.6.1 on default settings [21]. The best hit was an identical *C. perfringens* protein of 928 AAs (NCBI Accession No. WP_011590364) with a query coverage of 100% and an AA identity of 100%. This protein contained two types of conserved domains. A single von Willebrand Factor (vWF) type A domain of 142 AAs (NCBI accession no. smart00327) and three Cna protein B-type domains each of 63 AAs (NCBI accession no. pfam05738).

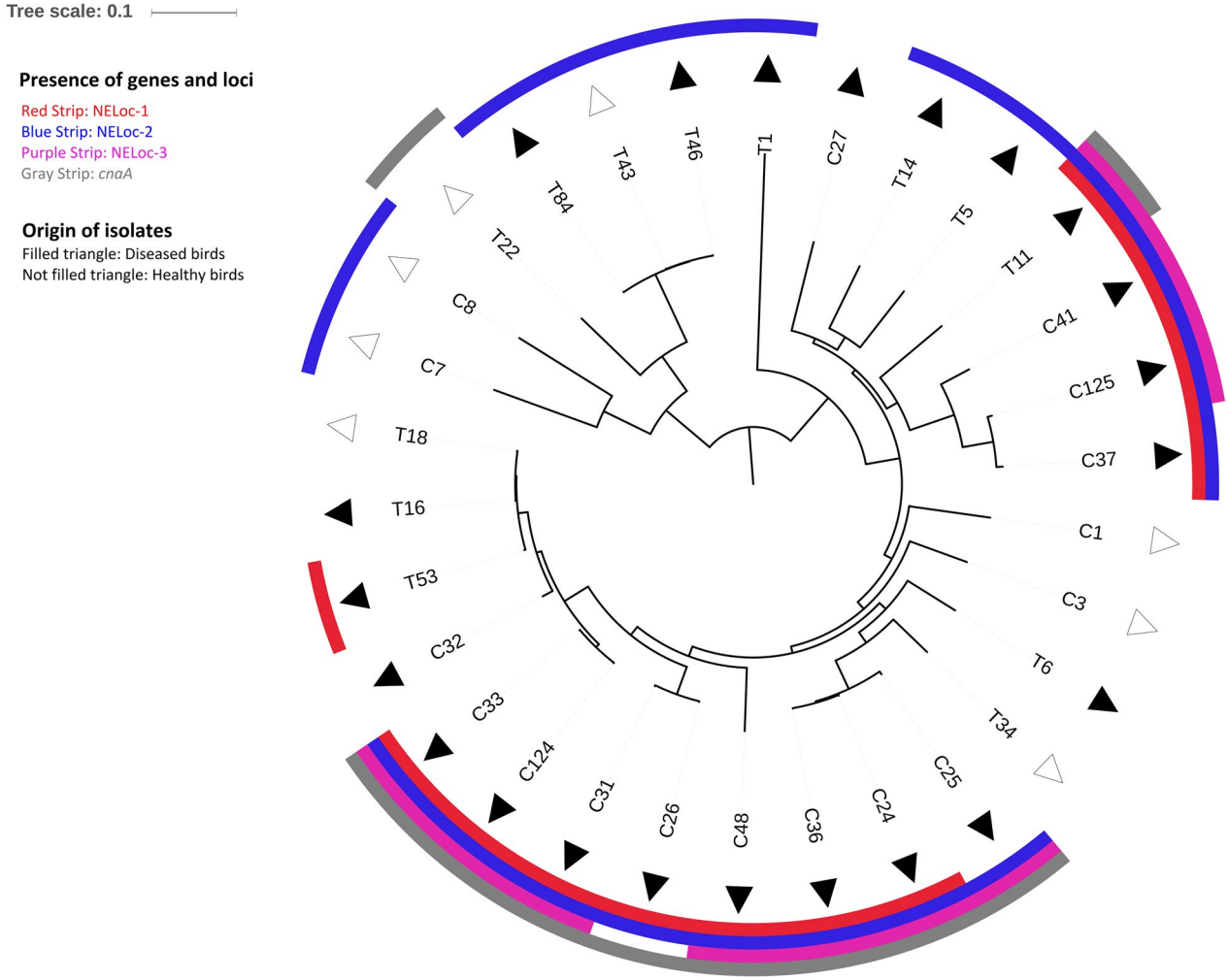


Fig. 1 Phylogenetic tree including the distribution of important virulence genes and loci. The maximum likelihood approximated tree based on 50,643 SNPs, shows that 30 isolates from healthy and diseased chickens (C) and the turkeys (T) did not cluster into representative groups. NELoc-1, -3 and *cnaA* were significantly present in NE isolates from chickens, whereas NELoc-2 was associated with NE isolates from both chickens and turkeys. Reference strain *C perfringens* ATCC 13124 is not included. The *scale* represents substitutions per site

Discussion

In this study we carried out genome analysis of *C. perfringens* isolates from healthy and NE afflicted turkeys and chickens. To our knowledge it is the first time *C. perfringens* isolates from turkeys have been whole-genome sequenced and made publicly available. The majority of the all isolates were found to be of unknown ST. This is probably because no *C. perfringens* isolates from the poultry environments investigated in this study, have previously been MLST and thus there are currently no defined MLST schemes that fits these isolates. A phylogenetic analysis confirmed that the majority of isolates were not closely related but constituted a relative diverse population of different genetic backgrounds.

NELoc-1 that carries *netB*, was found only in isolates from poultry with NE and primarily in chickens where 77% of the NE isolates were *netB*-positive (Table 1). Previous studies of chicken isolates showed that *netB* was predominantly present (>80%) among the NE isolates, but absent or rarely detected (<4%) in healthy chickens [4, 5]. Few studies of virulence genes in isolates from turkeys have been carried out. The turkey isolates analyzed in this study were from a previous study [8] where *netB* was found in 26% (14/55) of the NE isolates from turkeys, whereas all isolates from healthy turkeys were *netB*-negative. In another study [7], *netB* was not found in 42 NE isolates from turkeys. Like *netB*, *cnaA* was significantly present among NE isolates from chickens compared to a single NE isolate from a turkey (Table 1). Wade et al. [11]

found *cnaA* only in diseased chickens whereas another study [22] also identified *cnaA* in isolates from healthy chickens. Additionally, NELoc-2 and -3 were also primarily found in isolates from diseased chickens. NELoc-1, -2 and -3 were initially discovered in broilers with NE [10] which is why they also in this study were found highly represented among diseased chickens. In contrast, only NELoc-2 was present in the majority of the NE isolates from turkeys, suggesting that this pathogenicity locus may be associated with NE pathogenesis in turkeys (Fig. 1).

Interestingly, a putative collagen adhesion gene was discovered in all diseased turkeys and according to a suggested lettering system presented by Wade et al. [11], it has here been designated as *cnaD*. This gene encodes a vWF type A domain and three Cna protein B-type domains. These two types of domains have been shown to be involved in collagen binding [23]. A single Cna protein B-type domain, identical to those found in this study, is encoded by *cnaA* [11] and the vWF type A domain is also involved in collagen binding [24]. Thus, it may be suggested that these adhesion properties play an important role when *C. perfringens* strains attach to collagen in the intestine of turkeys and cause NE. This should be further investigated via in vitro studies. In a study of turkeys by Saita et al. [9], it was indicated that *C. perfringens* was not only associated with NE showing the same clinical and pathological changes as in broiler chickens, but also with other manifestations of intestinal disorders. These observations suggest that both the pathological manifestations and the pathogenesis of NE are different between turkeys and chickens.

To summarize, NELoc-1, -3, *netB* and *cnaA* were significantly associated with NE isolates from chickens, whereas only NELoc-2 was associated with NE isolates from both turkeys and chickens. Thus, *C. perfringens* virulence genes and loci in chickens with NE do not seem to be important to the same extent in diseased turkeys, suggesting that the NE pathogenesis is different in these two avian species. A putative collagen adhesion gene, *cnaD* was identified in all diseased turkeys and could potentially be of importance in regard to the NE pathogenesis.

Limitations

An increased sample size would have provided a more robust foundation for these findings. The expression of virulence genes could have been investigated through RNA sequencing which potentially could have further elucidated their importance during the NE pathogenesis. It should be mentioned that NE is a multifactorial disease and predisposing factors like feed composition, mycotoxins, temperature and hygiene stress do also play a considerable role during development of disease outbreaks on poultry farms [5, 25, 26]. Thus, it is not only the genetic profile of virulent *C. perfringens* strains that dictate the pathogenesis of NE.

Additional files

Additional file 1: Table S1. Thirty *C. perfringens* isolates from healthy and NE infected poultry. The pdf-file contains background information of each isolate including SRA accession numbers.

Additional file 2: Table S2. Descriptions of various *C. perfringens* genes. The pdf-file contains description of gene products, NCBI accession numbers and references of various *C. perfringens* genes.

Additional file 3: Table S3. Presence of NELoc-1 and -3 genes in the high prevalence groups. In this pdf-file, information regarding absence/presence of the ORFs that constitute NELoc-1 and -3 can be found. The table includes locus tags and gene product descriptions.

Additional file 4: Table S4. NELoc-2 genes identified among the *C. perfringens* isolates. In this pdf-file, information regarding absence/presence of the ORFs that constitute NELoc-2 can be found. The table includes locus tags and gene product descriptions.

Additional file 5. NT and AA sequence of the putative collagen adhesion gene *cnaD*.

Abbreviations

AA: amino acid; GW: Genomics Workbench; HPG: high prevalence group; LPG: low prevalence group; MW: Main Workbench; MLST: multi locus sequence typing; NE: necrotic enteritis; NELoc: necrotic enteritis locus; NT: nucleotide; ORF: open reading frame; ST: sequence type; vWF: von Willebrand Factor.

Authors' contributions

BL, KLN, KP, MS, TR and UL performed the experiments designed by KP, TR. KP, MS, PSA, TR and UL performed data analyses and interpretations. TR wrote the manuscript and all authors read and approved the final manuscript.

Author details

[1] National Veterinary Institute, Technical University of Denmark, Bülowsvej 27, 1870 Frederiksberg C, Denmark. [2] Department of Bacteria, Parasites and Fungi, Statens Serum Institut, Artillerivej 5, 2300 Copenhagen S, Denmark. [3] Veterinary Disease Biology, University of Copenhagen, Stigbøjlen 4, 1870 Frederiksberg C, Denmark.

Acknowledgements

We will like to thank Elvira Chapka and Mette Holm from Statens Serum Institut and Katja A. Kristensen from National Veterinary Institute for excellent laboratory work and advice.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The clinical *C. perfringens* isolates generated during and/or analyzed during the current study will be available from the corresponding author on reasonable request.

Ethics approval and consent to participate

The samples analyzed in this study were from dead poultry affected by NE. All farmers have volunteered to submit the dead animals for diagnostic examination. Thus, no ethical approval was required.

Publisher's Note

Received: 8 March 2017 Accepted: 3 July 2017
Published online: 11 July 2017

References

1. Songer JG. Clostridial enteric diseases of domestic animals. Clin Microbiol Rev. 1996;9:216–34.
2. Rood JI. Virulence genes of *Clostridium perfringens*. Annu Rev Microbiol. 1998;52:333–60.
3. McDonel JL. *Clostridium perfringens* toxins (type a, b, c, d, e). Pharm Ther. 1980;10:617–55.
4. Keyburn AL, Yan XX, Bannam TL, Van Immerseel F, Rood JI, Moore RJ. Association between avian necrotic enteritis and *Clostridium perfringens* strains expressing NetB toxin. Vet Res. 2010;41:21.
5. Keyburn AL, Boyce JD, Vaz P, Bannam TL, Ford ME, Parker D, et al. NetB, a new toxin that is associated with avian necrotic enteritis caused by *Clostridium perfringens*. PLoS Pathog. 2008;4:e26.
6. Wade B, Keyburn AL. The true cost of necrotic enteritis. World Poult. 2015;31:16–7.
7. Giovanardi D, Drigo I, De Vidi B, Agnoletti F, Viel L, Capello K, et al. Molecular characterization of *Clostridium perfringens* strains isolated from diseased turkeys in Italy. Avian Pathol. 2016;9457:1–17.
8. Lyhs U, Perko-Makela P, Kallio H, Brockmann A, Heinikainen S, Tuuri H, et al. Characterization of *Clostridium perfringens* isolates from healthy turkeys and from turkeys with necrotic enteritis. Poult Sci. 2013;92:1750–7.
9. Saita M, Bano L, Gallazzi DD. Pathogenicity markers of *Clostridium* spp. in commercial turkeys. Ital J Anim Sci. 2010;8:781–4.
10. Lepp D, Roxas B, Parreira VR, Marri PR, Rosey EL, Gong J, et al. Identification of novel pathogenicity loci in *Clostridium perfringens* strains that cause Avian necrotic enteritis. PLoS ONE. 2010;5:e10795.
11. Wade B, Keyburn AL, Seemann T, Rood JI, Moore RJ. Binding of *Clostridium perfringens* to collagen correlates with the ability to cause necrotic enteritis in chickens. Vet Microbiol. 2015;180:299–303.
12. Nauerby B, Pedersen K, Madsen M. Analysis by pulsed-field gel electrophoresis of the genetic diversity among *Clostridium perfringens* isolates from chickens. Vet Microbiol. 2003;94:257–66.
13. The National Center for Biotechnology Information. Sequence Read Archive. Rockville Pike: US National Library of Medicine; 2016. https://www.ncbi.nlm.nih.gov/sra. Accessed Sept 2016.
14. MyDbFinder v1.1. Lyngby: Center for Genomic Epidemiology; 2016. https://cge.cbs.dtu.dk/services/MyDbFinder. Accessed Jan 2016.
15. Chen L, Zheng D, Liu B, Yang J, Jin Q. VFDB 2016: hierarchical and refined dataset for big data analysis-10 years on. Nucleic Acids Res. 2016;44:D694–7.
16. Jolley KA, Chan M-S, Maiden MCJ. mlstdbNet—distributed multi-locus sequence typing (MLST) databases. BMC Bioinform. 2004;5:86.
17. Kaas RS, Leekitcharoenphon P, Aarestrup FM, Lund O. Solving the problem of comparing whole bacterial genomes across different sequencing platforms. PLoS ONE. 2014;9:1–8.
18. Letunic I, Bork P. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Res. 2011;39:W475–8.
19. Prokka Seemann T. Rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068–9.
20. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015;31:3691–3.
21. Altschul SF, Wootton JC, Gertz EM, Agarwala R, Morgulis A, Scha AA. Protein database searches using compositionally adjusted substitution matrices. FEBS. 2005;272:5101–9.
22. Lepp D, Gong J, Songer JG, Boerlin P, Parreira VR, Prescott JF. Identification of accessory genome regions in poultry *clostridium perfringens* isolates carrying the *netB* plasmid. J Bacteriol. 2013;195:1152–66.
23. Jost BH, Billington SJ, Trinh HT, Songer JG. Association of genes encoding beta2 toxin and a collagen binding protein in *Clostridium perfringens* isolates of porcine origin. Vet Microbiol. 2006;115:173–82.
24. Machha VR, Tischer A, Moon-Tasson L, Auton M. The von willebrand factor a1–collagen III interaction is independent of conformation and type 2 von willebrand disease phenotype. J Mol Biol. 2016;429:32–47.
25. Lovland A, Kaldhusdal M. Severely impaired production performance in broiler flocks with high incidence of *Clostridium perfringens*-associated hepatitis. Avian Pathol. 2001;30:73–81.
26. Moore RJ. Necrotic enteritis predisposing factors in broiler chickens. Avian Pathol. 2016;9457:1–22.

Table S1. 30 C. *perfringens* isolates from healthy and NE infected poultry

Isolate	Animal	State	Sample	Farm	Country/year	Reference	SRA acc. no.
C1	Chicken	Healthy	Intestine	1	Denmark/2001	(Nauerby et al., 2003)	SRR4448030
C3	Chicken	Healthy	Intestine	1	Denmark/2001	(Nauerby et al., 2003)	SRR4448031
C7	Chicken	Healthy	Intestine	2	Denmark/2001	(Nauerby et al., 2003)	SRR4448028
C8	Chicken	Healthy	Intestine	2	Denmark/2001	(Nauerby et al., 2003)	SRR4448029
C24	Chicken	Diseased	Intestine	3	Denmark/2001	(Nauerby et al., 2003)	SRR4457405
C25	Chicken	Diseased	Intestine	4	Denmark/2000	(Nauerby et al., 2003)	SRR4457404
C26	Chicken	Diseased	Intestine	5	Denmark/2000	(Nauerby et al., 2003)	SRR4457403
C27	Chicken	Diseased	Liver	6	Denmark/1997	(Nauerby et al., 2003)	SRR4457402
C31	Chicken	Diseased	Liver	7	Denmark/1998	(Nauerby et al., 2003)	SRR4457406
C32	Chicken	Diseased	Liver	8	Denmark/1998	(Nauerby et al., 2003)	SRR4457397
C33	Chicken	Diseased	Liver	9	Denmark/1998	(Nauerby et al., 2003)	SRR4448032
C36	Chicken	Diseased	Liver	10	Denmark/1998	(Nauerby et al., 2003)	SRR4457399
C37	Chicken	Diseased	Intestine	11	Denmark/1999	(Nauerby et al., 2003)	SRR4457398
C41	Chicken	Diseased	Intestine	10	Denmark/2002	(Nauerby et al., 2003)	SRR4457408
C48	Chicken	Diseased	Intestine	12	Denmark/2002	(Nauerby et al., 2003)	SRR4457407
C124	Chicken	Diseased	Liver	13	Denmark/1997	(Nauerby et al., 2003)	SRR4457401
C125	Chicken	Diseased	Intestine	14	Denmark/2000	(Nauerby et al., 2003)	SRR4457400
T1	Turkey	Diseased	Intestine	15	Finland/1998	(Lyhs et al., 2013)	SRR4434753
T5	Turkey	Diseased	Intestine	16	Finland/2005	(Lyhs et al., 2013)	SRR4434754
T6	Turkey	Diseased	Intestine	17	Finland/2005	(Lyhs et al., 2013)	SRR4434751
T11	Turkey	Diseased	Intestine	16	Finland/2006	(Lyhs et al., 2013)	SRR4434752
T14	Turkey	Diseased	Intestine	18	Finland/2006	(Lyhs et al., 2013)	SRR4434749
T16	Turkey	Diseased	Intestine	17	Finland/2008	(Lyhs et al., 2013)	SRR4434750
T18	Turkey	Healthy	Intestine	17	Finland/2009	(Lyhs et al., 2013)	SRR4434747
T22	Turkey	Healthy	Intestine	19	Finland/2009	(Lyhs et al., 2013)	SRR4434748
T34	Turkey	Healthy	Intestine	17	Finland/2009	(Lyhs et al., 2013)	SRR4434755
T43	Turkey	Healthy	Intestine	20	Finland/2009	(Lyhs et al., 2013)	SRR4434756
T46	Turkey	Diseased	Intestine	18	Finland/2010	(Lyhs et al., 2013)	SRR4434757
T53	Turkey	Diseased	Intestine	21	Finland/2010	(Lyhs et al., 2013)	SRR4434758
T84	Turkey	Diseased	Intestine	16	Finland/2011	(Lyhs et al., 2013)	SRR4434759

Table S2 Description of various *C. perfringens* genes

Gene	Location	Product	Activity	NCBI acc. no.	Reference
plc	chromosome	Alpha (α) toxin	Phosphlipase C	KP143661	-
cpb	plasmid	Beta (β) toxin	β pore-forming	KP064409	[1]
exp	plasmid	Epsilon (ε) toxin	β pore-forming	JX010451	[2]
iap	plasmid	Iota (ι) toxin	Actin ADP-ribosylation	NC_015712	[3]
cnaA	chromosome	Collagen adhesin	Adherence	KT749987	[4]
NELoc-1 33 ORFs Incl. *netB*	plasmid	toxins, adhesion factors, regulation/ mobilization proteins, enzymes	(NetB): β pore-forming, adhesion, regulating, mobilizing, enzymatic	JF837812	[5]/ [6]
NELoc-2 11 ORFs	chromosome	Regulation/ mobilization proteins, enzymes	regulating, mobilizing, enzymatic	JF837813	[5]/ [6]
NELoc-3 5 ORFs	plasmid	mobilization proteins, enzymes	mobilizing, enzymatic	NG_041400	[5]/ [6]

Genes investigated for presence among 30 *C. perfringens* isolates. Additionally, NCBI accession number, the genomic localization of the gene, gene product and activity are shown.

References

1. Theoret JR, Uzal FA, McClane BA. Identification and characterization of *Clostridium perfringens* beta toxin variants with differing trypsin sensitivity and in vitro cytotoxicity activity. Infect. Immun. 2015;83:1477–86.

2. Harkness JM, Li J, McClane BA. Identification of a lambda toxin-negative *Clostridium perfringens* strain that processes and activates epsilon prototoxin intracellularly. Anaerobe. Elsevier Ltd; 2012;18:546–52.

3. Miyamoto K, Yumine N, Mimura K, Nagahama M, Li J, McClane BA, et al. Identification of novel *clostridium perfringens* type E strains that carry an iota toxin plasmid with a functional enterotoxin gene. PLoS One. 2011;6.

4. Wade B, Keyburn AL, Seemann T, Rood JI, Moore RJ. Binding of *Clostridium perfringens* to collagen correlates with the ability to cause necrotic enteritis in chickens. Vet. Microbiol. Elsevier B.V.; 2015;180:299–303.

5. Lepp D, Gong J, Songer JG, Boerlin P, Parreira VR, Prescott JF. Identification of accessory genome regions in poultry *clostridium perfringens* isolates carrying the *netB* plasmid. J. Bacteriol. 2013;195:1152–66.

6. Lepp D, Roxas B, Parreira VR, Marri PR, Rosey EL, Gong J, et al. Identification of novel pathogenicity loci in *Clostridium perfringens* strains that cause Avian necrotic enteritis. PLoS One. 2010;5.

Table S3. Presence of NELoc-1 and -3 genes in the high prevalence groups

C24	C26	C31	C33	C36	C37	C41	C48	C124	C125	T11	T53	**NeLoc-1 (33 ORFs) Locus tag/gene product**
+	+	+	+	+	+	+	+	+	+	+	+	CP4_3442/transcriptional regulator, MarR family
+	+	+	+	+	+	+	+	+	+	+	+	CP4_3443/β-lactamase domain containing protein
+	+	+		+	+	+		+	+	+	+	CP4_3444/M protein trans-acting positive regulator (MGA)
+	+	+	+	+	+	+	+	+	+	+	+	CP4_3445/putative radical SAM domain containing protein
+	+	+		+	+	+		+	+	+	+	CP4_3446/putative internalin
+	+	+	+	+	+	+	+	+	+	+	+	CP4_3447/putative protein
+	+	+	+	+	+	+	+	+	+	+	+	CP4_3448/putative protein
+	+	+	+	+	+	+	+	+	+	+	+	CP4_3449/NE toxin B (NetB)
+	+	+	+	+	+	+	+	+	+	+	+	CP4_3450/ricin-type β-trefoil domain protein
+	+	+	+	+	+	+	+	+	+	+	+	CP4_3451/transposase for transposon
+	+	+	+	+	+	+	+	+	+	+	+	CP4_3452/hypothetical protein
+	+	+	+	+	+	+	+	+	+	+	+	CP4_3453/hypothetical protein
+	+	+	+	+	+	+	+	+	+	+	+	CP4_3454/chitinase B
+	+	+	+	+	+	+	+	+	+	+	+	CP4_3455/chitodextrinase
+	+	+	+	+	+	+	+	+	+	+	+	CP4_3457/CAAX amino terminal protease
+	+	+	+	+	+	+	+	+	+	+	+	CP4_3458/putative β-toxin
+	+	+	+	+	+	+	+	+	+	+	+	CP4_3459/putative β-toxin
+	+	+	+	+	+	+	+	+	+	+	+	CP4_3460/conserved hypothetical protein
+	+	+	+	+	+	+	+	+	+	+	+	CP4_3461/conserved hypothetical protein
+	+	+	+	+	+	+	+	+	+	+	+	CP4_3462/conserved hypothetical protein
+	+	+	+	+	+	+	+	+	+	+	+	CP4_3464/resolvase recombinase
+	+	+		+	+	+	+	+	+	+	+	CP4_3465/resolvase recombinase
+	+	+	+	+	+	+	+	+	+	+	+	CP4_3466/putative membrane protein
+	+	+	+	+	+	+	+	+	+	+	+	CP4_3468/F5/8 type C domain containing protein
+	+	+	+	+	+	+	+	+	+	+	+	CP4_3469/phosphodiesterase domain 2
+	+	+	+	+	+	+	+	+	+	+	+	CP4_3470/sortase family protein
+	+	+	+	+	+	+	+	+	+	+	+	CP4_3471/putative surface protein
+	+	+	+	+	+	+	+	+	+	+	+	CP4_3472/peptidoglycan bound protein
	+	+	+	+	+	+	+	+	+	+	+	CP4_3473/cell wall surface anchor protein
+	+	+	+	+	+	+	+	+	+	+	+	CP4_3474/signal peptidase I
+	+	+	+	+	+	+	+	+	+	+	+	CP4_3476/hypothetical protein
+	+	+	+	+	+	+	+	+	+	+	+	CP4_3477/swim zinc finger domain
+	+	+		+	+	+	+	+	+	+	+	CP4_3478/diguanylate cyclase
C24	**C26**	**C31**	**C33**	**C36**	**C37**	**C41**	**C48**	**C124**	**C125**	**T11**	**T53**	**NeLoc-3** (5 ORFs) Locus tag/gene product
+	+	+	+	+	+	+	+	+	+	+	+	CP4_3567/conserved hypothetical protein
+	+	+		+	+	+	+		+	+	+	CP4_3568/hypothetical protein
+	+	+	+	+	+	+	+	+	+	+	+	CP4_3569/resolvase/recombinase
+	+	+	+	+	+	+	+	+	+	+	+	CP4_3570/conserved hypothetical protein
+	+	+	+	+	+	+	+	+	+	+	+	CP4_3572/NADP-dependent 7-α-HSD

The locus tag and the gene product of the NELoc-1 and -3 associated genes are shown (Lepp et al., 2013). + indicates presence respectively, among the *C. perfringens* isolates. Only isolates with high prevalence of the NELoc-1 and -3 genes are presented.

Table S4. NELoc-2 genes identified among the *C. perfringens* isolates

C7	C8	C24	C26	C31	C33	C36	C37	C41	C48	C124	C125	T1	T5	T11	T43	NeLoc-2 (11 ORFs) Locus tag/gene product
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	CP4_0458/sigma factor SgiI
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	CP4_0459/conserved hypothetical protein
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	CP4_0460/putative VTC domain superfamily
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	CP4_0461/ putative tubulin/FtsZ, GTPase
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	CP4_0462/resolvase
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	CP4_0463/putative VTC domain superfamily
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	CP4_0464/ tubulin/FtsZ, GTPase
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	CP4_0465/ CotH protein
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	CP4_0466/conserved hypothetical protein
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	CP4_0467/putative heat repeat
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	CP4_0468 chitin synthase

The locus tag and the gene product of the NELoc-2 associated genes are shown (Lepp et al., 2013). + indicates presence of genes and NELoc-2 was fully conserved among all isolates from this table.

cnaD(2787NTs)

ATGAATAGTAAGGGTATAGATACCAAAAAATTAATGTCAATTATAAGCTTAATAATGACAGTTATATTTTTAAGTATTTTAT
TGCCTACTAATTTAACAAAAGCAGAAGAAAAGTCTGATAGCATGTCTGTTGAAAAAGTATTAAATTCAGAAATGTATAATA
ATTTCAATAATGAAATATTAAACAATAAAAGTAGTATTAATTACACTACTGATAATAATCAAGGTACATGGCCAGTAAATTG
GGAGTATGGAAATGTAAGCAATAAAAATAAAATAAACAAAAGTGTATATGGCAAGAATCAACCATTAGAATATAATGAAG
GATATTTAACAAAGAAAGCTTATACAACAGATGAAGATAATGTATTTGACATTAATCTAAAGATTCAGGGGAAAAAAAATC
AATCTTTAAAAAAAGATGTGGTTTTCCTTTTAGACAATTCAAATTCTATGACAACAAATAATCGTGCAATAAAGATTAAAGA
ACAAATTAAAAATGTTATGGATAAGCTAAATACTAATAATACACGTTATGCGCTTGTTACTTATGCCTCAACAATTTTAGAT
GGAAGGTATTATCATTTAATTGATAGATCTATAGGGGATAATAAATATACAGTTTATAAAGGTTATACAAGTAACCAGTGT
TATCTAAATTTTACTAGTAATATTCAAGAGATTTATAATAAAATACCTACTACTGTTCCAAATCAGAGAAATAATGGTTATGT
AGGGGGAACATTTACTCAAGAAGGATTATTGAAAGCAATAGAACTTTTAAAAAATAGTGATGCTGATGAAAAAATTATTA
TTCATCTTACTGATGGGTTACCAACATTTTCTTTTCTTTTAAAAGAGTTTGGAGGAAATGAAAAAGCTATTTTTGACTATAAC
ACTCAATATAATGGTATTGGTGTACGTGGATTTGGGACATCATACTTTTTTAATACTAAAACTCAAAAGCCGTATATATATT
CTAGAGAAGAAGTATATTCTGCTTTAAATCGTTCAATAAATAAAAATGAATCAATATGGAATAATGGTTTTCCAACTACTTT
AGAAGCAGAGAATATTAAAAAGGAAAATCCGGACATTAATATTTACACTATTGGGATAGAACTTAAAAAAGAAGTATATA
AATGGGATGATTATAGAAAATATTATAATGCTGAAGGTGTTGTTGAACTTCCAGAAATAAAAAAATTCTTAGAATCAATTT
CTTCTAGTCCTGCTGAGGCTTTTGTTAATGAAAATGTTGATGATATTGATGAGATTATTAATAAAATTATTGATAAAATAAA
GAATTCAATAAATGATGGTACTGTTATAGATCCTATGGGTGATATGGTTTATATTGTTAAAAATGGAGAATTTAATAATGA
GGATTATAAGTTAACGGCATCTAATAATAAGTTATTAGAAGGTGTAAAAGTAGGATATAACGAAAAAAATAGACAGATAG
TTCTTACAGGACTAAATTTAGGTGAAAATGAATGGGTTGAATTAAATTATAAAGTAAGATTAAATACAAGCAACCCTGACT
TTAATGGGGATTTTTGGTATCAAGCTAATAAAAGAACTGTTTTAAACCCTAATAATAAAGAACCAAATATTTTTCGTGACTT
TGTAATACCTTCTGTTAGTGGAAAAAGACCATCAATAGAAGTTAAATTAAAGAAGATATCAAGTGAAACATCAAAGCCATT
AGCTAATTCAGAATTTGAACTTTATAATTCTATAAATGAAAAATTAGGCTCTTTTACAACAAAAGAAAATGGGGAGGTAAG
TCTTGGGTATTTACCAGAAGGTGAGTATAAGTTAAAAGAAATAACCCCACCAAAAGGATATATTTTATCTAAAGACTTTATT
AATTTTAAGATTAATAATGGAAAGGCTATACAAGATGGAAATGAAGTAGAATTTATTACGGTAAGTAATAAAGTTAATAGT
ATATGTATAAAGAAAACTGATGATGCAAAGTTAGAAGCTGATGCTAAATTTTTAAGTGGAGCAAAATTTGAATTAAATAAA
GCTAATGATAAAAATTTCAAACCTTTAGTAAAAGAAACTGATGATAAGGGAGAAATAGAATTTAATGAGATTGAACCTGG
TACTTACTATTTAAAAGAAGTTCTTGCTCCTAATGGATATGAACAAATTAAGGAAGACATAGGACCAATTGTAGTTGATAA
TACTGGTGTAGTAACAATTCCTTGGGATAAATTAAAAAGCAACGATGTAGAAAAGTGGAATAATCAAGAGATTATTCGTAT
AAAAAATAAAAAGTTGAAATCGGCTGTATACATAGATAAAGTAGATGCAATAAATCAAGGAATAAAGTTAAGTGGAGCAA
AATTTTCACTTTATACTAATGATGAAAATTATAAAAATGATAAAAAGCTAGTAAGAAATGGTGTAAATTATTATTTAATAAG
TGAAAAAGTATCAAATTATGAAGGTAGAATTGAATGGGATAATCTTAATTCAGGACAAGAATATAAATATCTTATTCAGGA
GACTGAGGCACCAAAGGGATATACTGTAAGTGGAAAAGAAATATTATTCCATTTTAAAGATAATACTGTTGTTATAGATAA
TGAGAATGATGTAAAAGCACTTGCTAGTATAAATGGACAAGTTATTAGTATTAAAAATGCTAAAATATATAAGCTTCCGTC
ATCTGGCGGTATCGGAGTGTACCCTTTCTTACTTATAGGGACACTATGTATGGCTTTAAGTTTAATATATAGTTTTAATAGT
AAGGTTTTAAATAAAAGGAAATAA

cnaD(928AAs)

MNSKGIDTKKLMSIISLIMTVIFLSILLPTNLTKAEEKSDSMSVEKVLNSEMYNNFNNEILNNKSSINYTTDNNQGTWPVNWEYG
NVSNKNKINKSVYGKNQPLEYNEGYLTKKAYTTDEDNVFDINLKIQGKKNQSLKKDVVFLLDNSNSMTTNNRAIKIKEQIKNVM
DKLNTNNTRYALVTYASTILDGRYYHLIDRSIGDNKYTVYKGYTSNQCYLNFTSNIQEIYNKIPTTVPNQRNNGYVGGTFTQEGLL
KAIELLKNSDADEKIIIHLTDGLPTFSFLLKEFGGNEKAIFDYNTQYNGIGVRGFGTSYFFNTKTQKPYIYSREEVYSALNRSINKNESI
WNNGFPTTLEAENIKKENPDINIYTIGIELKKEVYKWDDYRKYYNAEGVVELPEIKKFLESISSSPAEAFVNENVDDIDEIINKIIDKIK
NSINDGTVIDPMGDMVYIVKNGEFNNEDYKLTASNNKLLEGVKVGYNEKNRQIVLTGLNLGENEWVELNYKVRLNTSNPDFN
GDFWYQANKRTVLNPNNKEPNIFRDFVIPSVSGKRPSIEVKLKKISSETSKPLANSEFELYNSINEKLGSFTTKENGEVSLGYLPEGE
YKLKEITPPKGYILSKDFINFKINNGKAIQDGNEVEFITVSNKVNSICIKKTDDAKLEADAKFLSGAKFELNKANDKNFKPLVKETDD
KGEIEFNEIEPGTYYLKEVLAPNGYEQIKEDIGPIVVDNTGVVTIPWDKLKSNDVEKWNNQEIIRIKNKKLKSAVYIDKVDAINQGI
KLSGAKFSLYTNDENYKNDKKLVRNGVNYYLISEKVSNYEGRIEWDNLNSGQEYKYLIQETEAPKGYTVSGKEILFHFKDNTVVID
NENDVKALASINGQVISIKNAKIYKLPSSGGIGVYPFLLIGTLCMALSLIYSFNSKVLNKRK*

Appendix IV

Paper IV

Ronco T, Stegger M, Pedersen K. 2017. Draft genome sequence of a sequence type 398 methicillin-resistant *Staphylococcus aureus* isolate from a danish dairy cow with mastitis. Genome Announc. 5, e00492-17.

AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements™

PROKARYOTES

Draft Genome Sequence of a Sequence Type 398 Methicillin-Resistant *Staphylococcus aureus* Isolate from a Danish Dairy Cow with Mastitis

Troels Ronco,[a] Marc Stegger,[b] Karl Pedersen[a]

National Veterinary Institute, Technical University of Denmark, Frederiksberg, Denmark[a]; Department of Bacteria, Parasites and Fungi, Statens Serum Institut, Copenhagen, Denmark[b]

ABSTRACT Livestock-associated (LA) methicillin-resistant *Staphylococcus aureus* (MRSA) strains of sequence type 398 (ST398) colonize both humans and various livestock species. In 2016, an ST398 LA-MRSA isolate (Sa52) was collected from a Danish dairy cow with mastitis, and here, we report the draft genome sequence of strain Sa52.

Received 19 April 2017 **Accepted** 20 April 2017 **Published** 8 June 2017

Citation Ronco T, Stegger M, Pedersen K. 2017. Draft genome sequence of a sequence type 398 methicillin-resistant *Staphylococcus aureus* isolate from a Danish dairy cow with mastitis. Genome Announc 5:e00492-17. https://doi.org/10.1128/genomeA.00492-17.

Address correspondence to Karl Pedersen, kape@vet.dtu.dk.

Staphylococcus aureus is a notorious opportunistic pathogen that infects both humans and animals and constitutes a major cause of bovine mastitis (1). In Denmark, livestock-associated (LA) methicillin-resistant *S. aureus* (MRSA) strains of sequence type 398 (ST398) primarily colonize pigs but are now also increasingly causing human infections (2, 3). The ST398 lineage has spread to various livestock species and has been detected in European dairy cattle with mastitis and in bulk-tank milk (1). However, there are hitherto no reports on clinical ST398 MRSA isolates from Danish dairy herds or Danish cattle in general. Here, we present a draft genome sequence of the first reported LA-MRSA isolate (Sa52) of ST398 collected in 2016 from a Danish dairy cow with mastitis.

The Nextera XT DNA kit (Illumina) was used to prepare libraries of fragmented DNA for paired-end sequencing, with an average read length of 2 × 150 bp, applying Illumina's NextSeq platform. The quality of raw reads was analyzed in FastQC 0.11.5 and *de novo* assembly performed in CLC bio's Genomics Workbench 10.0. The total assembly and N_{50} size were determined to be 2.8 Mb and 34.4 kb, respectively. An average coverage of 103×, 145 contigs, and a G+C content of 32.8% were obtained. Gene annotation was performed using NCBI's Prokaryotic Genome Automatic Annotation Pipeline (PGAAP) (4) and resulted in a total of 2,878 genes, of which 2,735 were protein-coding sequences (CDSs).

Multilocus sequence typing using PubMLST (5) showed that Sa52 was from ST398. Virulence and resistance genes were identified using ResFinder 2.1 (6) and Virulence Finder 1.5 (7), respectively, whereas the BLASTN 2.2.30+ (8) implementation in CLC bio's Genomics Workbench was applied for further identification of genomic content. The *mecA* gene and a wide range of other resistance genes (*blaZ*, *ermB*, *lnuB*, *norA*, *spc*, *tetK*, and *tetM*) were found. The staphylococcal cassette chromosome *mec* element (SCC*mec*) V(5C2&5)c (9) was entirely conserved, and PHASTER (10) identified an intact phage of 45.6 kb, StauST398-2 (GenBank accession no. NC_021323). Virulence genes (*aur*, *cap*, *fib*, *hla*, *hlb*, *hlg*, *icaD*, and *nuc*) that previously have been found in *S. aureus* strains from cattle with mastitis and in bulk-tank milk were detected (11, 12). Last, a single *hsdM* gene from a type 1 restriction-modification (RM) system was identified via Restriction-Modification Finder 1.1 (http://cge.cbs.dtu.dk). Since no *hsdS* genes were

present, it indicated that Sa52 did not carry any functional RM systems, which means that foreign DNA is easily taken up (13).

The ST398 LA-MRSA lineage has disseminated to both humans and various livestock species, and here, we report the first case, to our knowledge, of a clinical isolate of this lineage collected from a Danish dairy cow with mastitis. For this reason, we recommend monitoring of clinical infections in cattle and/or bulk-tank milk samples for MRSA to follow any potential development in the distribution and frequency of LA-MRSA.

Accession number(s). This whole-genome shotgun project has been deposited in DDBJ/ENA/GenBank under the accession number NBNF00000000. The version described in this paper is the first version.

ACKNOWLEDGMENTS

We thank Kári K. Mouritsen from the National Veterinary Institute and Mette T. Christiansen from SSI for great DNA sequencing work.

REFERENCES

1. Holmes MA, Zadoks RN. 2011. Methicillin resistant *S. aureus* in human and bovine mastitis. J Mammary Gland Biol Neoplasia 16:373–382. https://doi.org/10.1007/s10911-011-9237-x.
2. Agersø Y, Hasman H, Cavaco LM, Pedersen K, Aarestrup FM. 2012. Study of methicillin resistant *Staphylococcus aureus* (MRSA) in Danish pigs at slaughter and in imported retail meat reveals a novel MRSA type in slaughter pigs. Vet Microbiol 157:246–250. https://doi.org/10.1016/j.vetmic.2011.12.023.
3. Larsen J, Petersen A, Sørum M, Stegger M, Van Alphen L, Valentiner-Branth P, Knudsen LK, Larsen LS, Feingold B, Price LB, Andersen PS, Larsen AR, Skov RL. 2015. Meticillin-resistant *Staphylococcus aureus* CC398 is an increasing cause of disease in people with no livestock contact in Denmark, 1999 to 2011. Euro Surveill 20:pii=30021. http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=21245.
4. Angiuoli SV, Gussman A, Klimke W, Cochrane G, Field D, Garrity GM, Kodira CD, Kyrpides N, Madupu R, Markowitz V, Tatusova T, Thomson N, White O. 2008. Toward an online repository of Standard Operating Procedures (SOPs) for (Meta)genomic annotation. OMICS 12:137–141. https://doi.org/10.1089/omi.2008.0017.
5. Jolley KA, Maiden MCJ. 2010. BIGSdb: scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics 11:595. https://doi.org/10.1186/1471-2105-11-595.
6. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV. 2012. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother 67:2640–2644. https://doi.org/10.1093/jac/dks261.
7. Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, Nielsen EM, Aarestrup FM. 2014. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. J Clin Microbiol 52:1501–1510. https://doi.org/10.1128/JCM.03617-13.
8. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402. https://doi.org/10.1093/nar/25.17.3389.
9. Vandendriessche S, Vanderhaeghen W, Larsen J, de Mendonça R, Hallin M, Butaye P, Hermans K, Haesebrouck F, Denis O. 2014. High genetic diversity of methicillin-susceptible *Staphylococcus aureus* (MSSA) from humans and animals on livestock farms and presence of SCC*mec* remnant DNA in MSSA CC398. J Antimicrob Chemother 69:355–362. https://doi.org/10.1093/jac/dkt366.
10. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res 44:W16–W21. https://doi.org/10.1093/nar/gkw387.
11. Bardiau M, Caplin J, Detilleux J, Graber H, Moroni P, Taminiau B, Mainil JG. 2016. Existence of two groups of *Staphylococcus aureus* strains isolated from bovine mastitis based on biofilm formation, intracellular survival, capsular profile and *agr*-typing. Vet Microbiol 185:1–6. https://doi.org/10.1016/j.vetmic.2016.01.003.
12. Xu J, Tan X, Zhang X, Xia X, Sun H. 2015. The diversities of staphylococcal species, virulence and antibiotic resistance genes in the subclinical mastitis milk from a single Chinese cow herd. Microb Pathog 88:29–38. https://doi.org/10.1016/j.micpath.2015.08.004.
13. Murray NE. 2000. Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). Microbiol Mol Biol Rev 64:412–434. https://doi.org/10.1128/MMBR.64.2.412-434.2000.

Appendix V

Paper V

Ronco T, Klaas IC, Stegger M, Svennesen L, Astrup LB, Farre M, Pedersen K. 2018. Genomic investigation of *Staphylococcus aureus* isolates from bulk tank milk and dairy cows with clinical mastitis. Vet. Microbiol. 215, 35-42.

Veterinary Microbiology 215 (2018) 35–42

Contents lists available at ScienceDirect

Veterinary Microbiology

journal homepage: www.elsevier.com/locate/vetmic

Genomic investigation of *Staphylococcus aureus* isolates from bulk tank milk and dairy cows with clinical mastitis

Troels Ronco[a,*], Ilka C. Klaas[b], Marc Stegger[c], Line Svennesen[b], Lærke B. Astrup[a], Michael Farre[d], Karl Pedersen[a]

[a] National Veterinary Institute, Technical University of Denmark, Kemitorvet Build. 204, 2800 Kgs. Lyngby, Denmark
[b] Section for Production, Nutrition and Health, University of Copenhagen, Grønnegårdsvej 2, 1870 Fdr. C, Denmark
[c] Department of Bacteria, Parasites and Fungi, Statens Serum Institut, Artillerivej 5, 2300 Copenhagen S, Denmark
[d] SEGES Livestock Innovation, Agro Food Park 15, 8200 Aarhus N, Denmark

ARTICLE INFO

Keywords:
Genomic analysis
S. aureus
Clinical mastitis
Bulk tank milk
Virulence and resistance genes

ABSTRACT

Staphylococcus aureus is one of the most common pathogens that cause mastitis in dairy cows. Various subtypes, virulence genes and mobile genetic elements have been associated with isolates from bulk tank milk and clinical mastitis. So far, no Danish cattle associated *S. aureus* isolates have been whole-genome sequenced and further analyzed. Thus, the main objective was to investigate the population structure and genomic content of isolates from bulk tank milk and clinical mastitis, using whole-genome sequencing. This may reveal the origin of strains that cause clinical mastitis.

S. aureus isolates from bulk tank milk (n = 94) and clinical mastitis (n = 63) were collected from 91 and 24 different farms, respectively and whole-genome sequenced. The genomic content was analyzed and a phylogenetic tree based on single nucleotide polymorphisms was constructed.

In general, the isolates from both bulk tank milk and clinical mastitis were of similar genetic background. This suggests that dairy cows are natural carriers of the *S. aureus* subtypes that cause clinical mastitis if the right conditions are present and that a broad range of subtypes cause mastitis. A phylogenetic cluster that mostly consisted of ST151 isolates carried three mobile genetic elements that were primarily found in this group. The prevalence of resistance genes was generally low. However, the first ST398 methicillin resistant *S. aureus* isolate from a Danish dairy cow with clinical mastitis was detected.

1. Introduction

Staphylococcus aureus is an opportunistic pathogen that may cause severe infections in both humans and livestock and is a major cause of mastitis in dairy cows (Holmes and Zadoks, 2011; Agersø et al., 2012; Larsen et al., 2015). Bovine mastitis results in reduced animal welfare, milk quality and milk production which is the reason for remarkable economic losses worldwide (Halasa et al., 2007; Haran et al., 2012; Barkema et al., 2009). A variety of different sequence types (STs) (ST97, 126 133, 151, 479 and 771) (Holmes and Zadoks, 2011; Zadoks et al., 2011) and *spa*-types (t518, t519, t524 t528, t529 and t543) have been associated with bovine mastitis and cattle worldwide (Hasman et al., 2010; Ikawaty et al., 2009; Sakwinska et al., 2011). Previous studies have shown that few types of strains belonging to specific genotypes are successful at causing persistent mastitis and strain RF122 (ST151) has been reported as one of the most common clone types involved in clinical mastitis (CM) (Kapur et al., 1995; Reinoso et al., 2004; Haveri et al., 2005; Fritzgerald et al., 1997). This strain carries various mobile genetic elements (MGEs) that contain virulence genes and other types of genes related to host adaptation. Most of these genes are found within specific types of MGEs such as *S. aureus* pathogenicity islands (SaPIs), phages and genomic islands (Herron-Olson et al., 2007). In general, various types of virulence genes have been detected in clinical and subclinical mastitis isolates and in bulk tank milk (BTM). These virulence factors are involved in: Host colonization (*cap, clfA/B, cna, fib* and *sak*), toxin production (*tst*, *sea-j*, *hla/b/g*, *lukD/E/FM/S*, *etA/B*) and biofilm formation (*icaD, fnbB*) (Fueyo et al., 2005; Bardiau et al., 2016; Xu et al., 2015; Fournier et al., 2008; Yamada et al., 2005). Many of these virulence genes encode toxins that are also harmful to humans. For example, staphylococcal enterotoxins (encoded by *se* genes) cause food poisoning, the toxic shock syndrome toxin-1 (encoded by *tst*) causes toxic shock syndrome and leukocidins (encoded by *lukD/E/FS*) are involved in various types of clinical infections (Asao et al., 2003; Umeda et al., 2017; Deurenberg et al., 2005; Lina et al., 1999).

* Corresponding author.
E-mail address: troro@vet.dtu.dk (T. Ronco).

https://doi.org/10.1016/j.vetmic.2018.01.003
Received 8 November 2017; Received in revised form 8 January 2018; Accepted 9 January 2018

Methicillin resistant *S. aureus* (MRSA) belonging to ST398 has been observed among bovine mastitis isolates across the globe but has not disseminated among Danish herds of dairy cows (Holmes and Zadoks, 2011; Zadoks et al., 2011). However, in Denmark this lineage has primarily been found in pigs and is now an increasing cause of human infections (Agersø et al., 2012; Larsen et al., 2015). Previously, various studies of Danish *S. aureus* isolates associated with bovine mastitis have been carried out (Katholm et al., 2012; Aarestrup et al., 1995a; Larsen et al., 2000a, 2002, 2000b) but no Danish isolates from BTM and CM have so far been whole-genome sequenced. Thus, the main objective of this study was to investigate the genomic content and population structure of Danish *S. aureus* isolates from BTM and CM, using whole-genome sequencing. A further objective was to investigate possible differences between the BTM and the CM isolates.

2. Materials and methods

2.1. S. aureus isolates

In 2016, CM isolates (n = 63) were all sampled from different cows on 24 different Danish farms. The aseptic foremilk samples were collected from dairy cows with CM according to the National Mastitis Council's guidelines. Samples of mastitis secrete or plates with growth were submitted to the Danish Veterinary Institute for *S. aureus* verification using Matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF).

Isolates from BTM (n = 94) were sampled from 91 different Danish dairy farms. The farms were selected based on the yearly BTM samples taken under a surveillance program for *Streptococcus agalactiae* as previously described (Katholm et al., 2012). Samples were analyzed with the Mastit4 real-time PCR test (DNA diagnostic A/S, Risskov, Denmark) at an analytic laboratory (Eurofins, Vejen, Denmark). Based on the PCR test result, 100 herds with the lowest Ct-value (ranging from 21 to 27) were selected and samples submitted to the Danish Veterinary Institute. The BTM samples were cultured by streaking 10 µl on blood agar (Columbia agar base (Oxoid, CM0331, Hampshire, UK) supplemented with 5% calf blood) and on *S. aureus* selective ager (SA Select, bioMérieux, Marcy-l'Étoile, France). Colonies suspected for being *S. aureus* were further sub-cultured and verified as *S. aureus* using MALDI-TOF.

Both BTM and CM isolates were sampled from different farms distributed in all parts of the country.

2.2. DNA purification and whole-genome sequencing

S. aureus colonies were grown overnight on blood agar at 37 °C and single colonies were cultured in 5 ml trypticase soy broth (Becton-Dickinson and Company, Franklin Lakes, USA) under the same conditions. DNA was purified using a Maxwell 16 LEV Blood DNA Kit (Promega, Madison, USA) according to manufacturer's instructions, with an additional lysis-phase including 200 µg/ml lysostaphin per sample (Sigma-Aldrich, St. Louis, USA). Subsequently, a Nextera XT kit (Illumina, San Diego, USA) was used for building DNA libraries according to manufacturer's instructions. The DNA libraries were paired-end sequenced applying Illumina's NextSeq platform with a read length on 2 × 151 bp. The Illumina sequence reads have been deposited in NCBI's short read archive with the study accession no. SRP119902.

2.3. De novo assembly and subtyping

The quality of the Illumina raw reads was analyzed in FastQC 0.11.5 and bases of low quality were trimmed in CLC bio's Genomics Workbench (GW) v10.0 (CLCbio's, Aarhus, Denmark) using default settings. Subsequently, *de novo* assembly was performed in CLC bio's GW on default settings and a minimum contig size of 500 nt. MLST was performed at PubMLST (Jolley et al., 2004) and MLST v1.8 (Larsen et al., 2012) whereas *spa*-types were determined using spaTyper v1.0 (Bartels et al., 2014).

2.4. Identification of genomic content

The isolates were investigated for all resistance and virulence genes in *de novo* assembled contigs using ResFinder v2.1 (Zankari et al., 2012) and VirulenceFinder v1.5 (Joensen et al., 2014), respectively. Furthermore, the prevalence of four specific virulence genes (*fib*, *hla*, *icaD* and *nuc*) from strain Sa52 (Ronco et al., 2017) was investigated. Subsequently, the genes were identified in the assemblies using the BLASTN (Altschul et al., 1997) implementation in CLC bio's Main Workbench (MW) v7.7.3 and in general, if genes were located on > 1 contig CLC bio's MW was used to identify these. The presence of ORFs that belonged to eight different MGEs (Tables S1–S8) was investigated using CLC bio's GW.

2.5. Statistics

Statistical analyses were performed using GraphPad Prism v5.02 (GraphPad Software Inc., San Diego, USA). Differences in the presence of STs, *spa*-types and virulence/resistance genes between BTM and CM isolates were investigated applying a Chi-square test for independence. In cases of ≤5 observations, a Fisher's exact test was used. The confidence interval was 95% and the difference considered significant when $P < 0.05$.

2.6. Identification of single nucleotide polymorphisms

To investigate the relationship between the 157 isolates single nucleotide polymorphisms (SNPs) were identified using CSI Phylogeny v1.4 (Kaas et al., 2014) with *S. aureus* strain ED133 as reference chromosome (accession no. NC_017337). The SNPs were identified with a quality of ≥30, a minimum depth of ≥10× and a distance between SNPs of ≥10. Subsequently, a phylogenetic tree was visualized using iTOL v3.6.1 (Letunic and Bork, 2011).

3. Results

3.1. MLST and spa-typing

All isolates had an average sequencing depth of > 50 fold except a single that had 47 fold. Statistical analyses showed no significant differences in distributions of STs or *spa*-types between BTM and CM isolates, except for ST1 and ST97 that were significantly associated with CM isolates (Table 1). Thirty different STs were found and 12 of these were new and subsequently registered at PubMLST (Jolley et al., 2004). Among BTM isolates 27 different STs were observed whereas 15 were found among the CM isolates. The most prevalent of the new STs were ST3891 and ST3897 found in 17% (27/157) and 5% (8/157) of all isolates, respectively (Table 1). Of the remaining STs, the prevalence of the six most commonly found (ST50, 71, 97, 133, 151 and 479) ranged 5–19% with ST151 as the most prevalent (Table 1). Among all isolates, 15 different *spa*-types were observed. However, 24 BTM and 15 CM isolates were identified as being of unknown *spa*-type. The prevalence of the six most often observed *spa*-types (t519, t524, t528, t529, t543 and t1403) ranged 5–27%, with t529 as the most prevalent (Table 1).

3.2. Resistance and virulence genes

In general, all genes were identified with thresholds of 90% nucleotide identity and 90% coverage of the query sequence length. Statistical analyses showed no significant differences in distributions of resistance genes between BTM and CM isolates. Ten different antibiotic resistance genes were observed. The *norA* gene was found in all isolates except a single one, whereas the second most prevalent resistance gene, *blaZ* was observed in 17% (27/157) of the isolates. Only 9% (14/157)

Table 1
Prevalence of STs and *spa*-types among *S. aureus* isolates.

ST (%)	CM (n = 63)	BTM (n = 94)	*P*-value	*spa*-type (%)	CM (n = 63)	BTM (n = 94)	*P*-value
151 (19)	11	19	0.6672	**t529** (27)	14	29	0.2347
3891* (17)	8	19	0. 2213	**t519** (10)	6	10	0.8210
133 (9)	9	5	0.0838	**t1403** (10)	9	7	0.1650
97 (6)	8	2	*0.0151*	**t528** (6)	4	5	1.0000
479 (6)	4	5	1.0000	**t524** (5)	2	6	0.4768
50 (5)	6	2	0.0608	**t543** (5)	4	3	0.4395
71 (5)	2	6	0.4768	t2873 (4)	1	5	0.4027
3897* (5)	4	4	0.7146	t518 (3)	2	3	1.0000
504 (4)	2	4	1.0000	t693 (1)	1	0	0.4013
1 (3)	4	0	*0.0245*	t948 (1)	2	0	0.1595
1380 (3)	0	5	0.0831	t1200 (1)	1	0	0.4013
398 (2)	1	2	1.0000	t2207 (1)	1	0	0.4013
705 (2)	1	2	1.0000	t4911 (1)	0	1	1.0000
7 (1)	0	1	1.0000	t7652 (1)	0	1	1.0000
8 (1)	1	0	0.4013	t7750 (1)	1	0	0.4013
9 (1)	1	1	1.0000	Unk (25)	15	24	0.8066
15 (1)	0	1	1.0000				
132 (1)	0	1	1.0000				
706 (1)	0	1	1.0000				
2423 (1)	0	2	0.5164				
3892* (1)	0	1	1.0000				
3896* (1)	1	0	0.4013				
3898* (1)	0	2	0.5164				
3899* (1)	0	2	0.5164				
3900* (1)	0	2	0.5164				
4361* (1)	0	1	1.0000				
4362* (1)	0	1	1.0000				
4363* (1)	0	1	1.0000				
4364* (1)	0	1	1.0000				
4365* (1)	0	1	1.0000				

The table shows the prevalence of 30 STs and 15 *spa*-types of *S. aureus* identified among 63 clinical mastitis (CM) isolates and 94 isolates from bulk tank milk (BTM). The eight most prevalent STs and six most prevalent *spa*-types are in bold whereas 12 new STs are marked with an asterisk (*). Statistical differences in distributions of STs and *spa*-types between CM and BTM isolates were investigated using statistical tests and significant *P*-values are in italic. Unk: Unknown.

Table 2
Prevalence of virulence and resistance genes among *S. aureus* isolates.

Virulence genes (%)	CM (n = 63)	BTM (n = 94)	*P*-value	Resistance genes (%)	CM (n = 63)	BTM (n = 94)	*P*-value
aur (100)	63	94	1.0000	*norA* (99)	62	94	0.4013
hla (100)	63	94	1.0000	*blaZ* (17)	13	14	0.3501
hlb (99)	63	93	1.0000	*tetM* (3)	1	3	0.6495
hlgB (99)	63	93	1.0000	*dfrG* (2)	0	3	0.2746
hlgC (96)	61	89	0.7027	*ermB* (1)	1	0	0.4013
fib (96)	59	92	0.2196	*lnuA* (1)	0	1	1.0000
nuc (95)	61	88	0.4768	*lnuB* (1)	1	1	1.0000
icaD (94)	61	87	0.3162	*mecA* (1)	1	0	0.4013
hlgA (94)	58	89	0.5236	*tetK* (1)	1	0	0.4013
splA (92)	60	85	0.3638	*vgaA* (1)	0	1	1.0000
splB (92)	60	85	0.3638				
lukD (89)	55	85	0.7296				
lukE (81)	49	78	0.4165				
seu (69)	36	72	*0.0099*				
sem (68)	34	72	*0.0030*				
sen (68)	36	70	*0.0231*				
seo (66)	33	71	*0.0026*				
sei (66)	34	69	*0.0120*				
seg (45)	23	47	0.0955				
splE (16)	15	10	*0.0270*				
sec (5)	2	5	0.7027				
sel (5)	2	5	0.7027				
tst (5)	2	5	0.7027				
scn (3)	3	1	0.3029				
seh (3)	4	0	*0.0245*				
seq (2)	3	0	0.0628				
sak (2)	3	0	0.0628				
sek (2)	3	0	0.0628				
sea/sep (2)	3	0	0.0628				

The table shows the prevalence of 29 virulence genes and 10 resistance genes identified among 63 clinical mastitis (CM) isolates and 94 isolates from bulk tank milk (BTM). Virulence genes were divided into three groups according to prevalence among all isolates: Group 1 (genes found in ≥81%), Group 2 (genes found in 45–69%) and Group 3 (genes found in 2–16%). Statistical differences in distributions of virulence and resistance genes between CM and BTM isolates were investigated using statistical tests and significant *P*-values are in italics.

of all isolates carried other types of resistance genes than *blaZ* and *norA* (Table 2). Altogether, 82% (129/157) of all isolates carried no other resistance genes than *norA* (data not shown). Among 62 of the 63 CM isolates only *blaZ* and *norA* were found whereas a single ST398 isolate carried a wide range of resistance genes (*blaZ*, *ermB, lnuB*, *mecA*, *norA*, *tetK*, and *tetM*) (data not shown)

Twenty-nine different virulence genes were identified and they could be divided into three groups according to prevalence among all 157 isolates. One group consisted of the 13 most prevalent genes (*aur*, *hla*, *hlb hlgB*, *splA/B*, *lukD/E*, *hlgA/C*, *nuc*, *fib* and *icaD*) found in ≥81% of all isolates. In the second group, the prevalence of six enterotoxin genes (*seg, sei*, *sem*, *sen*, *seo* and *seu*) ranged 45–69%, whereas the prevalence of the ten remaining genes (*sec*, *seh*, *sek*, *sel*, *seq*, *sea/sep*, *splE*, *tst*, *scn* and *sak*) in the third group ranged 2–16% (Table 2). According to statistical analyses five enterotoxin genes (*sei*, *sem*, *sen*, *seo* and *seu*) were significantly associated with BTM isolates whereas a serine protease gene (*spl*E) and an enterotoxin gene (*seh*) were significantly associated with CM isolates (Table 2). When looking at the combination of virulence genes found among isolates within the eight most prevalent STs, no ST97 isolates carried any enterotoxin genes whereas 1/8 of the ST71 and 1/14 of the ST133 isolates carried a single enterotoxin gene, *sei* (Table S9). The only types of enterotoxin genes that were found among the eight most prevalent STs were the six most prevalent types (*seg, sei*, *sem*, *sen*, *seo* and *seu*) (Table S9).

3.3. Identification of MGEs

In seven different MGEs, open reading frames (ORFs) were primarily identified with thresholds of > 80% nucleotide identity and 90% coverage of the query sequence length. In some cases the ORFs were identified with thresholds of > 70% nucleotide identity and a MGE was only considered present if ≥80% of its ORFs were present. Our analyses showed that some types of MGEs were primarily present among isolates with closely related genetic background (Table 3). The three MGEs; *φ*12bov, *υ*SaBov and *φ*SaBov-*υ*-Sa*βφ* were only identified among a group of closely related isolates that primarily belonged to ST151, except for *υ*SaBov that was also found in a single ST7 isolate. Isolates in this group that did not belong to ST151 were either single locus variants (SLVs) (ST3899, 3900 and 705) or double locus variants (DLVs) (ST504) of ST151. Additionally, SaPIbov1 and SaPIbov-*υ*Saα were primarily found in six ST504 isolates but also in a single ST705 and a single ST71 isolate. In contrast, SaPIbov4 and SaPIbov5 were found among isolates from many different STs that were not closely related and none of the 157 isolates carried SaPIbov2 (Table 3). A visual overview of the distribution of MGEs among all isolates can be found in Fig. 1.

3.4. Phylogenetic analysis

The SNP analysis included 38782 variant positions and 67.8% of the reference chromosome was covered by all isolates. In general, the phylogenetic tree showed that both CM and BTM isolates clustered together into groups of identical or closely related STs (Fig. 1). The largest cluster in the tree consisted primarily of 30 ST151 isolates whereas the remaining isolates were either SLVs (ST3899, 3900 and 705) or DLVs (ST504) of ST151. The second largest cluster consisted of 41 isolates whereof eight belonged to ST50 and the post prevalent of the remaining new STs were 26 ST3891 isolates (SLVs of ST50). The third largest cluster consisted of nine ST479 isolates, eight ST3897 isolates and five ST1380 isolates. Furthermore, smaller clusters primarily including isolates that belonged to ST133, ST97 and ST71 were present (Fig. 1).

4. Discussion

Here, we carried out whole-genome sequencing to investigate the population structure and genomic content of 157 Danish *S. aureus* isolates from BTM and dairy cows with CM. To our knowledge it is the first time that such types of Danish isolates have been whole-genome sequenced and made publicly available. Statistical analyses showed no significant differences in the distribution of *spa*-types or STs between the two isolate groups except for ST97 and ST1 that were significantly more associated with CM isolates compared to BTM isolates (Table 1). Only four of all isolates belonged to ST1 and therefore it is difficult to conclude further on this finding. Isolates from BTM samples may originate from subclinical infected quarters, but also from extra-mammary sites such as teat skin, teat canal and the cow environment or from milking staff (Haveri et al., 2008). This could be the reason for finding a more diverse composition of STs in BTM samples (27 different STs) compared to CM samples (15 different STs). Furthermore, the BTM isolates were collected from farms that had shown an increased concentration of *S. aureus* in BTM (PCR Ct-values: 21–27). A study suggests that Ct-values < 32 very likely can be interpreted as reflecting *S. aureus* intra-mammary infections (Mahmmod et al., 2017). Therefore, it is likely that the isolates from BTM were partly from cows with subclinical mastitis which is a mild form of mastitis that requires further testing to be recognized by the farmer. It has been described that a high strain heterogeneity can be interpreted as evidence of environmental mastitis (Klaas and Zadoks, 2017). Thus, the BTM isolates that showed increased strain heterogeneity compared to the CM isolates, could be associated with environmental mastitis

The phylogenetic analysis showed that the BTM isolates clustered together with CM isolates of identical or closely related STs. A large cluster of isolates that primarily belonged to ST151 was observed and the majority of the isolates in this cluster carried three MGEs (*φ*12bov, *υ*SaBov and *φ*SaBov-*υ*-Sa*βφ*) found in strain RF122 (Herron-Olson et al., 2007). These three MGEs were exclusively found in this cluster except for a single ST7 isolate that also carried one of these MGEs (Fig. 1). Strain RF122 belonged to ST151 and has been reported to be a commonly observed mastitis causing clone type (Fritzgerald et al., 1997). The MGEs originating from RF122 contained various virulence genes and therefore it has been suggested the these MGEs play an important role regarding the CM pathogenesis and successful adaption to dairy cows (Herron-Olson et al., 2007). The MGE, *φ*SaBov-*υ*Sa*βφ* carries leucocidin genes (*lukE/D*), serine proteases (*spl*C/*E/F*) and enterotoxin genes (sec/g/i/m/n/o) whereas *υ*SaBov carries streptolycin genes (Tables S6 and S7). The streptolycin genes encode leucocidin homologs that originates from *Streptococcus pyogenes* (Herron-Olson et al., 2007) and many of the virulence genes carried by *φ*SaBov-*υ*Sa*βφ* have previously been found among mastitis isolates (Fueyo et al., 2005; Bardiau et al., 2016; Fournier et al., 2008; Kot et al., 2016). Additionally, these three MGEs (*φ*12bov, *υ*SaBov and *φ*SaBov-*υ*-Sa*βφ*) contain many hypothetical and phage related genes (Tables S6–S8) that encode proteins of unknown functions and further studies could reveal which potential role they play. In the same way, SaPIbov4 and 5 carried many hypothetical genes that could be further investigated (Tables S3 and S4). Additionally, both SaPIbov4 and 5 carried a von willebrand factor-binding gene (*vwb*), that potentially could be of importance to the mastitis pathogenesis (Viana et al., 2010). Both statistical and phylogenetic analyses showed that the BTM and CM isolates in general were of identical genetic background. This correspond to other studies (Boss et al., 2016; Conceicão, 2017; Jørgensen et al., 2005) which found that STs and *spa*-types that were often associated with bovine mastitis are also present in healthy cows and BTM. These findings indicate that dairy cows are natural carriers of *S. aureus* subtypes that can cause CM, for example if the cows appear immunocompromised combined with poor milking practices and hygiene etc.

Some of the most prevalent STs (ST97, 133, 151 and 479) found in this study have previously been associated with bovine mastitis (Holmes and Zadoks, 2011; Zadoks et al., 2011; Boss et al., 2016) whereas others (ST50 and 71) have been related to healthy cows and BTM (Smith et al., 2005; Hata et al., 2010). In addition, the two most

Table 3
Presence of MGEs among *S. aureus* isolates.

MGE	Strain/CC	No. of isolates (CM)	No. of isolates (BTM)	ST	Reference
SaPIbov1	RF122/151	–	1	705	Fitzgerald et al. (2001)
SaPIbov1	RF122/151	2	4	504	Fitzgerald et al. (2001)
SaPIbov1	RF122/151	–	1	71	Fitzgerald et al. (2001)
SaPIbov2	V329/126	–	–	–	Cucarella et al. (2001)
SaPIbov-υSaα	RF122/151	–	1	705	Herron-Olson et al. (2007)
SaPIbov-υSaα	RF122/151	2	4	504	Herron-Olson et al. (2007)
SaPIbov-υSaα	RF122/151	–	1	71	Herron-Olson et al. (2007)
SaPIbov4	BA4/97	–	1	4365*	Viana et al. (2010)
SaPIbov4	BA4/97	–	1	4363*	Viana et al. (2010)
SaPIbov4	BA4/97	–	1	3898*	Viana et al. (2010)
SaPIbov4	BA4/97	1	–	3896*	Viana et al. (2010)
SaPIbov4	BA4/97	1	7	3891*	Viana et al. (2010)
SaPIbov4	BA4/97	–	1	706	Viana et al. (2010)
SaPIbov4	BA4/97	8	1	97	Viana et al. (2010)
SaPIbov4	BA4/97	2	6	71	Viana et al. (2010)
SaPIbov4	BA4/97	6	1	50	Viana et al. (2010)
SaPIbov5	JP5338/-	–	1	4365*	Viana et al. (2010)
SaPIbov5	JP5338/-	–	1	4363*	Viana et al. (2010)
SaPIbov5	JP5338/-	–	1	3898*	Viana et al. (2010)
SaPIbov5	JP5338/-	1	–	3896*	Viana et al. (2010)
SaPIbov5	JP5338/-	1	6	3891*	Viana et al. (2010)
SaPIbov5	JP5338/-	2	6	71	Viana et al. (2010)
SaPIbov5	JP5338/-	6	1	50	Viana et al. (2010)
υSaBov	RF122/151	–	2	3899*	Herron-Olson et al. (2007)
υSaBov	RF122/151	–	2	3900*	Herron-Olson et al. (2007)
υSaBov	RF122/151	1	2	705	Herron-Olson et al. (2007)
υSaBov	RF122/151	2	4	504	Herron-Olson et al. (2007)
υSaBov	RF122/151	11	19	151	Herron-Olson et al. (2007)
υSaBov	RF122/151	–	1	7	Herron-Olson et al. (2007)
φ12bov	RF122/151	–	2	3899*	Herron-Olson et al. (2007)
φ12bov	RF122/151	–	2	3900*	Herron-Olson et al. (2007)
φ12bov	RF122/151	1	2	705	Herron-Olson et al. (2007)
φ12bov	RF122/151	2	4	504	Herron-Olson et al. (2007)
φ12bov	RF122/151	11	17	151	Herron-Olson et al. (2007)
φSaBov-υ-Saβφ	RF122/151	–	2	3899*	Herron-Olson et al. (2007)
φSaBov-υ-Saβφ	RF122/151	–	1	3900*	Herron-Olson et al. (2007)
φSaBov-υ-Saβφ	RF122/151	1	1	705	Herron-Olson et al. (2007)
φSaBov-υ-Saβφ	RF122/151	2	3	504	Herron-Olson et al. (2007)
φSaBov-υ-Saβφ	RF122/151	6	15	151	Herron-Olson et al. (2007)

The table shows seven mobile genetic islands (MGEs) identified among 94 isolates from bulk tank milk (BTM) and 63 isolates from clinical mastitis (CM) and the STs for these isolates are shown. SaPI bov2 was not identified in any of the isolates. The number of open reading frames (ORFs) associated with each MGEs are shown together with the strains and their CCs, from where the MGEswere initially found. References for each MGE are shown and new STs are marked with an asterisk (*). Unk: Unknown.

prevalent of the 12 new STs ST3891 and ST3897 were SLVs of ST50 and ST479, respectively. The most prevalent *spa*-type was t529 and observed in 27% (43/157) of the isolates, followed by t1403 and t519 that were both found in 10% (16/157) of the isolates. These three *spa*-types have all been associated with bovine mastitis (Ikawaty et al., 2009; Sakwinska et al., 2011; Boss et al., 2016; Johler et al., 2011) but also healthy cows (Hasman et al., 2010). All 30 ST151 isolates belonged to *spa*-type t529 which correspond to a previous study (Sakwinska et al., 2011). Many isolates (39/157) were identified as being of unknown *spa*-type using spaTyper (Bartels et al., 2014). The main reason for this was that the *spa* genes were located on > 1 contig and therefore not all repeats were identified by spaTyper (Bartels et al., 2014). Assembly and sequencing error could also explain why the order of the *spa* repeats was not determined correctly. However, it was beyond the scope of this study to register new *spa*-types according to guidelines found at the Ridom SapServer (http://www.spaserver.ridom.de/).

In general, the prevalence of resistance genes was low which correspond to a previous study of Danish mastitis isolates where 81% of 105 isolates were susceptible to 11 antibiotics (Aarestrup et al., 1995a). However, the first ST398 livestock-associated (LA) MRSA isolate (Sa52) from a Danish dairy cow with CM was discovered. The fact that strain Sa52 carried many other resistance genes than the rest of the CM isolates that only carried *blaZ* or *norA*, indicates that it has been transmitted to the dairy cow from an environment with a different selective pressure in regard to antibiotics. Interestingly, previous studies suggest the ST398 lineage has the ability to efficiently jump between humans and livestock and cause severe humans infections (Larsen et al., 2015; Price et al., 2012). Additionally, it has been suggested that human epidemic MRSA clones originate from isolates within CC97 that have jumped from cows to humans (Spoor et al., 2013). This makes it important to further monitor cattle herds to avoid potential problems regarding LA-MRSA with zoonotic potential, even though the transmission of for example ST398 strains between humans and dairy cows seems to occur less frequently (Sakwinska et al., 2011; Boss et al., 2016). LA-MRSA isolates belonging to ST398 have primarily been associated with pigs but strain Sa52 was sampled on a farm where no pig farming had taken place (Larsen et al., 2015; Price et al., 2012). Thus, it is unlikely that this strain was directly related to pig farming but it could have been transmitted by a visitor or farm worker carrying the clone. Seventeen% of all isolates carried *blaZ* which correspond to a previous Danish study where 17% of 105 isolates produced beta-lactamases (Aarestrup et al., 1995a). Another previous study found no beta-lactamase genes among 70 CM isolates (Artursson et al., 2016). Remarkably, *norA* was found in all isolates except a single one. This gene encodes a multidrug drug resistance efflux pump that mediates resistance to quinolones and a variety of other antiseptic compounds (Santos Costa et al., 2015; Kaatz and Seo, 1995). Since fluorquinolones are not used to treat Danish dairy cows for CM (https://www.foedevarestyrelsen.dk/Leksikon/Sider/VetStat.aspx) the presence of this gene must be driven by other factors. It may be suggested that *norA*

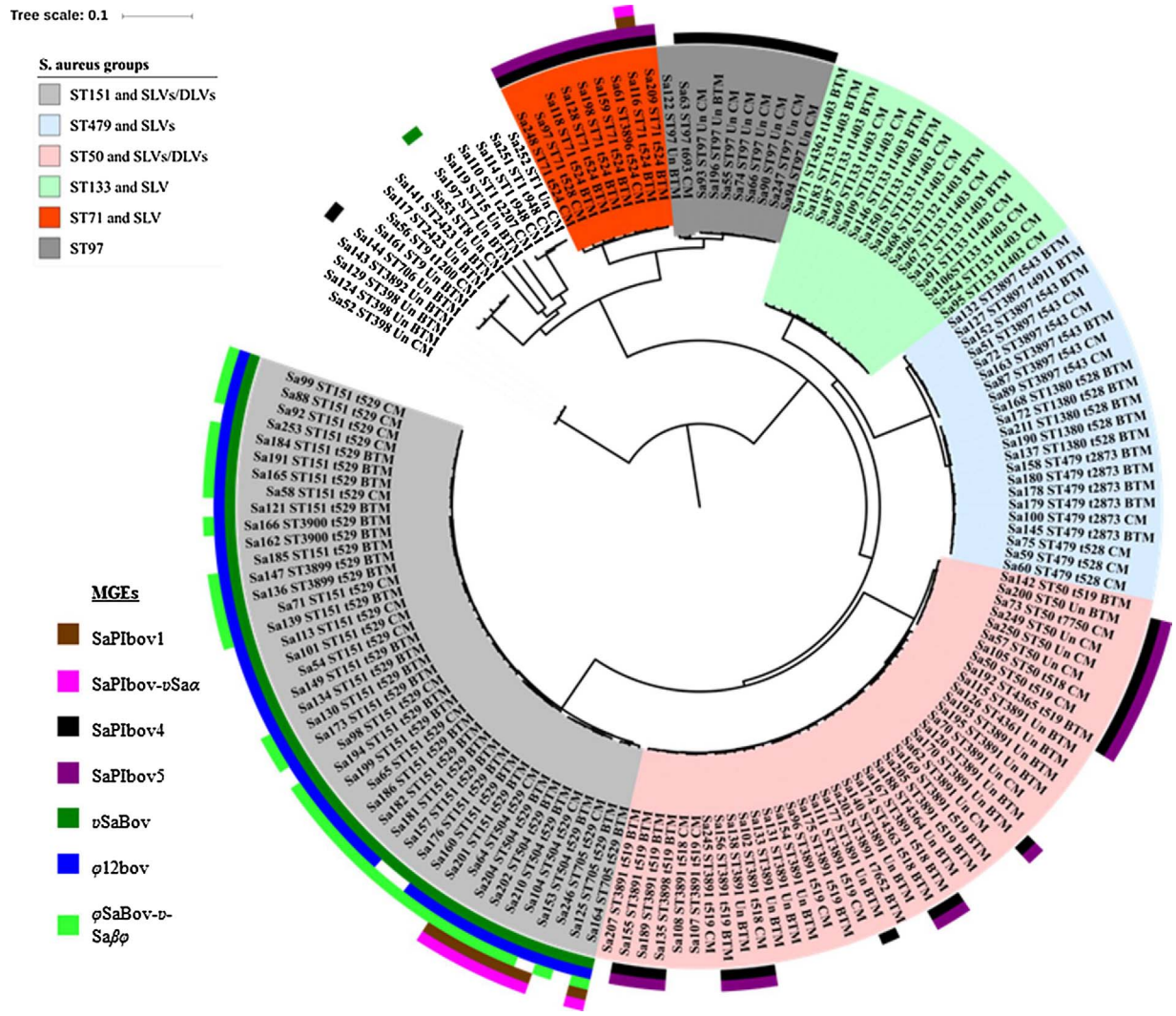


Fig. 1. Phylogenetic tree including the distribution of MGEs.
The phylogenetic tree is based on 38782 SNPs and *S. aureus* strain ED133 was used as reference chromosome. It shows that the 157 *S. aureus* isolates from bulk tank milk (BTM) and clinical mastitis (CM) cluster together into groups of closely related STs. Additionally, the distribution of seven different mobile genetic elements (MGEs) is shown as colored strips. Three different MGEs (*υ*SaBov and *φ*SaBov-*υ*-Sa*βφ*) were only found in a cluster that primarily consisted of ST151 isolates (marked in light gray). This cluster included single locus variants (SLVs) (ST3899, 3900 and 705) and double locus variants (DLVs) (ST504) of ST151. Only a single ST7 isolate outside this cluster carried *υ*SaBov. In contrast, SaPIbov4 and SaPIbov5 were found among various STs that were not all closely related.

caused resistance to antiseptic compounds used to increase the hygiene in the Danish dairy industry.

Six enterotoxin genes (*seg*, *sei*, *sem*, *sen*, *seo* and *seu*) were found more frequently than the rest of the enterotoxin genes and were additionally the only types found among isolates from the eight most prevalent STs (Table S9). Five of these genes were significantly more associated with BTM isolates compared to CM isolates. Previously, many of these genes have been found in *S. aureus* isolates from bovine mastitis (Fueyo et al., 2005; Xu et al., 2015; Fournier et al., 2008; Kot et al., 2016). However, the role of enterotoxins in the mastitis pathogenesis is not clear and studies indicate that they are not essential (Larsen et al., 2002, 2000b). Enterotoxins are heat-stabile and may therefore be found in various dairy products such as milk after heat treatment (Jørgensen et al., 2005; Hennekinne et al., 2012). Previously, enterotoxins have been reported to be associated with staphylococcal food poisoning caused by cow milk or other dairy products and even with mastitis in humans (Asao et al., 2003; Jørgensen et al., 2005; Hennekinne et al., 2012; Franck et al., 2017). Interestingly, isolates from the highly prevalent ST97 and ST133 that have been found to be strongly associated with CM (Holmes and Zadoks, 2011; Zadoks et al., 2011), carried almost no enterotoxins genes (Table S9). Concordantly, a previous PCR investigation of 106 Danish *S. aureus* isolates from subclinical and CM showed that none of the isolates carried any enterotoxin genes (Aarestrup et al., 1995b). Currently, more than 20 types of enterotoxin genes have been identified (Hennekinne et al., 2012) but in this study only 12 types were investigated. The VirulenceFinder database (Joensen et al., 2014) contains only 18 enterotoxin genes. Thus, it is possible that the isolates carried other novel enterotoxin genes, or genes related to mastitis pathogenesis and host adaption, than those included in this study (Table 2).Furthermore, a high proportion (≥81%) of all isolates carried the leukotoxin encoding genes *lukD* and *lukE*. These genes are often found in isolates associated with bovine mastitis (Fueyo et al., 2005; Bardiau et al., 2016) but have also been detected in clinical isolates from humans (Yoong and Torres, 2014).

5. Conclusion

In summary, both statistical and phylogenetic analyses showed that isolates from BTM and CM generally were of similar genetic background. This suggests that dairy cows can be natural carriers of, or subclinically infected with, *S. aureus* subtypes that can cause CM if the right conditions are present. A large cluster primarily consisting of

ST151 isolates carried three MGEs that were almost only found in this group and probably are involved in host adaption and the mastitis pathogenesis. A high proportion of all isolates carried leukotoxin genes and other toxin genes whereas five enterotoxin genes were significantly more associated with BTM isolates compared to CM isolates. Thus, both BTM and CM isolates carried genes that encode toxins that are harmful to humans. The prevalence of resistance genes was in general low but the first ST398 LA-MRSA isolate from a Danish dairy cow with CM was detected. Further surveillance of the Danish dairy cows is therefore important in order to avoid dissemination of zoonotic pathogens.

Conflicts of interest statement

All authors declare no conflicts of interest.

Acknowledgements

This project was supported by grants from Promilleafgiftsfonden and the Danish Milk Levy Fund (Mælkeafgiftsfonden). This project was also a part of the “STOPMAST” project. Furthermore, we would like to thank Mette T. Christiansen from Statens Serum Institut and Kári K. Mouritsen from the Technical University of Denmark for technical assistance.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.vetmic.2018.01.003.

References

Aarestrup, Wegener, H.C., Rosdahl, V.T., 1995a. Evaluation of phenotypic and genotypic methods for epidemiological typing of *Staphylococcus aureus* isolates from bovine mastitis in Denmark. Vet. Microbiol. 45, 139–150. http://dx.doi.org/10.1016/0378-1135(95)00043-A.

Aarestrup, Andersen, J.K., Jensen, N.E., 1995b. Lack of Staphylococcal enterotoxin production among strains of *Staphylococcus aureus* from bovine mastitis in Denmark. Acta Vet. Scand. 36, 273–275.

Agersø, Y., Hasman, H., Cavaco, L.M., Pedersen, K., Aarestrup, F.M., 2012. Study of methicillin resistant *Staphylococcus aureus* (MRSA) in Danish pigs at slaughter and in imported retail meat reveals a novel MRSA type in slaughter pigs. Vet. Microbiol. 157, 246–250. http://dx.doi.org/10.1016/j.vetmic.2011.12.023.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST:a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402. http://dx.doi.org/10.1093/nar/25.17.3389.

Artursson, K., Söderlund, R., Liu, L., Monecke, S., Schelin, J., 2016. Genotyping of *Staphylococcus aureus* in bovine mastitis and correlation to phenotypic characteristics. Vet. Microbiol. 193, 156–161. http://dx.doi.org/10.1016/j.vetmic.2016.08.012.

Asao, T., Kumeda, Y., Kawai, T., Shibata, T., Oda, H., Haruki, K., Nakazawa, H., Kozaki, S., 2003. An extensive outbreak of staphylococcal food poisoning due to low-fat milk in Japan: estimation of enterotoxin A in the incriminated milk and powdered skim milk. Epidemiol. Infect. 130, 33–40. http://dx.doi.org/10.1017/S0950268802007951.

Bardiau, M., Caplin, J., Detilleux, J., Graber, H., Moroni, P., Taminiau, B., Mainil, J.G., 2016. Existence of two groups of *Staphylococcus aureus* strains isolated from bovine mastitis based on biofilm formation, intracellular survival, capsular profile and *agr*-typing. Vet. Microbiol. 185, 1–6. http://dx.doi.org/10.1016/j.vetmic.2016.01.003.

Barkema, H.W., Green, M.J., Bradley, A.J., Zadoks, R.N., 2009. The role of contagious disease in udder health. J. Dairy Sci. 92, 4717–4729. http://dx.doi.org/10.3168/jds.2009-2347.

Bartels, M.D., Petersen, A., Worning, P., Nielsen, J.B., Larner-Svensson, H., Johansen, H.K., Andersen, L.P., Jarløv, J.O., Boye, K., Larsen, A.R., Westh, H., 2014. Comparing whole-genome sequencing with sanger sequencing for *spa* typing of methicillin-resistant *Staphylococcus aureus*. J. Clin. Microbiol. 52, 4305–4308. http://dx.doi.org/10.1128/JCM.01979-14.

Boss, R., Cosandey, A., Luini, M., Artursson, K., Bardiau, M., Al, E., 2016. Bovine *Staphylococcus aureus*: subtyping, evolution, and zoonotic transfer. J. Dairy Sci. 515–528.

Conceicão, T., 2017. Healthy bovines as reservoirs of major pathogenic lineages of *Staphylococcus aureus* in Portugal. Microb. Drug Resist. 1–7. http://dx.doi.org/10.1089/mdr.2017.0074.

Cucarella, C., Solano, C., Valle, J., Amorena, B., Lasa, I.G.O., Penade, R., 2001. Bap, a *Staphylococcus aureus* surface protein involved in biofilm formation. J. Bacteriol. 183, 2888–2896. http://dx.doi.org/10.1128/JB.183.9.2888.

Deurenberg, R.H., Nieuwenhuis, R.F., Driessen, C., London, N., Stassen, F.R., Tiel, F.H., Van Stobberingh, E.E., Vink, C., 2005. The prevalence of the *Staphylococcus aureus tst* gene among community- and hospital-acquired strains and isolates from Wegener's Granulomatosis patients. FEMS Microbiol. Lett. 245, 185–189. http://dx.doi.org/10.1016/j.femsle.2005.03.002.

Fitzgerald, J.R., Monday, S.R., Foster, T.J., Bohach, G.A., Hartigan, P.J., Meaney, W.J., Smyth, C.J., 2001. Characterization of a putative pathogenicity island from bovine *Staphylococcus aureus* encoding multiple superantigens. J. Bacteriol. 183, 63–70. http://dx.doi.org/10.1128/JB.183.1.63.

Fournier, C., Kuhnert, P., Frey, J., Miserez, R., Kirchhofer, M., Kaufmann, T., Steiner, A., Graber, H.U., 2008. Bovine *Staphylococcus aureus*: association of virulence genes, genotypes and clinical outcome. Res. Vet. Sci. 85, 439–448. http://dx.doi.org/10.1016/j.rvsc.2008.01.010.

Franck, K.T., Gumpert, H., Olesen, B., Larsen, A.R., Petersen, A., Bangsborg, J., Albertsen, P., Westh, H., 2017. *Staphylococcal aureus* enterotoxin C and enterotoxin-like l associated with post-partum mastitis. Front. Microbiol. 8, 1–5. http://dx.doi.org/10.3389/fmicb.2017.00173.

Fritzgerald, J.R., Meaney, W.J., Hartigan, P.J., Smyth, C.J., Kapur, V., 1997. Fine-structure molecular epidemiological analysis of *Staphylococcus aureus* recovered from cows. Epidemiol. Infect. 261–269.

Fueyo, J.M., Mendoza, M.C., Rodicio, M.R., Muñiz, J., Alvarez, M.A., Martín, M.C., 2005. Cytotoxin and pyrogenic toxin superantigen gene profiles of *Staphylococcus aureus* associated with subclinical mastitis in dairy cows and relationships with macrorestriction genomic profiles. J. Clin. Microbiol. 43, 1278–1284. http://dx.doi.org/10.1128/JCM.43.3.1278-1284.2005.

Halasa, T., Huijps, K., Østerås, O., Hogeveen, H., 2007. Economic effects of bovine mastitis and mastitis management: a review. Vet. Q. 29, 18–31. http://dx.doi.org/10.1080/01652176.2007.9695224.

Haran, K.P., Godden, S.M., Boxrud, D., Jawahir, S., Bender, J.B., Sreevatsan, S., 2012. Prevalence and characterization of *Staphylococcus aureus*, including methicillin-resistant *Staphylococcus aureus*, isolated from bulk tank milk from Minnesota dairy farms. J. Clin. Microbiol. 50, 688–695. http://dx.doi.org/10.1128/JCM.05214-11.

Hasman, H., Moodley, A., Guardabassi, L., Stegger, M., Skov, R.L., Aarestrup, F.M., 2010. *spa* type distribution in *Staphylococcus aureus* originating from pigs, cattle and poultry. Vet. Microbiol. 141, 326–331. http://dx.doi.org/10.1016/j.vetmic.2009.09.025.

Hata, E., Katsuda, K., Kobayashi, H., Uchida, I., Tanaka, K., Eguchi, M., 2010. Genetic variation among *Staphylococcus aureus* strains from bovine milk and their relevance to methicillin-resistant isolates from humans. J. Clin. Microbiol. 48, 2130–2139. http://dx.doi.org/10.1128/JCM.01940-09.

Haveri, M., Taponen, S., Pyo, S., 2005. Bacterial genotype affects the manifestation and persistence of bovine *Staphylococcus aureus* intramammary infection. J. Clin. Microbiol. 43, 959–961. http://dx.doi.org/10.1128/JCM.43.2.959.

Haveri, M., Hovinen, M., Roslo, A., Pyo, S., 2008. Molecular types and genetic profiles of *Staphylococcus aureus* strains isolated from bovine intramammary infections and extramammary sites. J. Clin. Microbiol. 46, 3728–3735. http://dx.doi.org/10.1128/JCM.00769-08.

Hennekinne, J., Buyser, M., De Dragacci, S., 2012. *Staphylococcus aureus* and its food poisoning toxins: characterization and outbreak investigation. FEMS Microb. Rev. 36, 815–836. http://dx.doi.org/10.1111/j.1574-6976.2011.00311.x.

Herron-Olson, L., Fitzgerald, J.R., Musser, J.M., Kapur, V., 2007. Molecular correlates of host specialization in *Staphylococcus aureus*. PLoS One 2 (10), e1120.

Holmes, M.A., Zadoks, R.N., 2011. Methicillin resistant *S. aureus* in human and bovine mastitis. J. Mammary Gland Biol. Neoplasia 16, 373–382. http://dx.doi.org/10.1007/s10911-011-9237-x.

Ikawaty, R., Brouwer, E.C., Jansen, M.D., Duijkeren, E., Van Mevius, D., Verhoef, J., Fluit, A.C., 2009. Characterization of dutch *Staphylococcus aureus* from bovine mastitis using a multiple locus variable number tandem repeat analysis. Vet. Microbiol. 136, 277–284. http://dx.doi.org/10.1016/j.vetmic.2008.10.034.

Jørgensen, H.J., Mørk, T., Rørvik, L.M., 2005. The occurrence of *Staphylococcus aureus* on a farm with small-scale production of raw milk cheese. J. Dairy Sci. 88, 3810–3817. http://dx.doi.org/10.3168/jds.S0022-0302(05)73066-6.

Joensen, K.G., Scheutz, F., Lund, O., Hasman, H., Kaas, R.S., Nielsen, E.M., Aarestrup, F.M., 2014. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. J. Clin. Microbiol. 52, 1501–1510. http://dx.doi.org/10.1128/JCM.03617-13.

Johler, S., Layer, F., Stephan, R., 2011. Comparison of virulence and antibiotic resistance genes of food poisoning outbreak isolates of *Staphylococcus aureus* with isolates obtained from bovine mastitis milk and pig carcasses. J. Food Prot. 74, 1852–1859. http://dx.doi.org/10.4315/0362-028X.JFP-11-192.

Jolley, K.A., Chan, M.-S., Maiden, M.C.J., 2004. mlstdbNet – distributed multi-locus sequence typing (MLST) databases. BMC Bioinformatics 5, 86. http://dx.doi.org/10.1186/1471-2105-5-86.

Kaas, R.S., Leekitcharoenphon, P., Aarestrup, F.M., Lund, O., 2014. Solving the problem of comparing whole bacterial genomes across different sequencing platforms. PLoS One 9, 1–8. http://dx.doi.org/10.1371/journal.pone.0104984.

Kaatz, G.W., Seo, S.M., 1995. Inducible NorA-mediated multidrug resistance in *Staphylococcus aureus*. Antimicrob. Agents Chemother. 39, 2650–2655.

Kapur, V., Sischo, W.M., Greer, R.S., Whittam, T.S., Musser, J.M., 1995. Molecular population genetic analysis of *staphylococcus aureus* recovered from cows. J. Clin. Microbiol. 33, 376–380.

Katholm, J., Bennedsgaard, T.W., Koskinen, M.T., Rattenborg, E., 2012. Quality of bulk tank milk samples from Danish dairy herds based on real-time polymerase chain reaction identification of mastitis pathogens. J. Dairy Sci. 95, 5702–5708. http://dx.doi.org/10.3168/jds.2011-5307.

Klaas, I.C., Zadoks, R.N., 2017. An update on environmental mastitis: challenging perceptions. Transbound. Emerg. Dis. 1–20. http://dx.doi.org/10.1111/tbed.12704.

Kot, B., Szweda, P., Frankowska-Maciejewska, A., Piechota, M., Wolska, K., 2016. Virulence gene profiles in *Staphylococcus aureus* isolated from cows with subclinical mastitis in eastern Poland. J. Dairy Res. 83, 228–235. http://dx.doi.org/10.1017/S002202991600008X.

Larsen, H.D., Sloth, K.H., Elsberg, C., Enevoldsen, C., 2000a. The dynamics of *Staphylococcus aureus* intramammary infection in nine Danish dairy herds. Vet. Microbiol. 71, 89–101.

Larsen, Huda, A., Eriksen, N.H.R., Jensen, N.E., 2000b. Differences between Danish bovine and human *Staphylococcus aureus* isolates in possession of superantigens. Vet. Microbiol. 76, 153–162.

Larsen, H.D., Aarestrup, F.M., Jensen, N.E., 2002. Geographical variation in the presence of genes encoding superantigenic exotoxins and beta-hemolysin among *Staphylococcus aureus* isolated from bovine mastitis in Europe and USA. Vet. Microbiol. 85, 61–67.

Larsen, J., Petersen, A., Sørum, M., Stegger, M., Van Alphen, L., Valentiner-Branth, P., Knudsen, L.K., Larsen, L.S., Feingold, B., Price, L.B., Andersen, P.S., Larsen, A.R., Skov, R.L., 2015. Meticillin-resistant *staphylococcus aureus* CC398 is an increasing cause of disease in people with no livestock contact in Denmark, 1999 to 2011. Eurosurveillance 20. http://dx.doi.org/10.2807/1560-7917.ES.2015.20.37.30021.

Larsen, M.V., Cosentino, S., Rasmussen, S., Friis, C., Hasman, H., Marvig, R.L., Jelsbak, L., Sicheritz-Ponten, T., Ussery, D.W., Aarestrup, F.M., Lund, O., 2012. Multilocus sequence typing of total-genome-sequenced bacteria. J. Clin. Microbiol. 50, 1355–1361. http://dx.doi.org/10.1128/JCM.06094-11.

Letunic, I., Bork, P., 2011. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Res. 39, W475–W478. http://dx.doi.org/10.1093/nar/gkr201.

Lina, G., Piémont, Y., Godail-gamot, F., Peter, M., 1999. Involvement of Panton-valentine leukocidin –producing *Staphylococcus aureus* in primary skin infections and pneumonia. Clin. Inectious Dis. 1128–1132.

Mahmmod, Y.S., Klaas, I.C., Enevoldsen, C., 2017. DNA carryover in milk samples from routine milk recording used for PCR-based diagnosis of bovine *Staphylococcus aureus* mastitis. J. Dairy Sci. 100, 5709–5716. http://dx.doi.org/10.3168/jds.2016-12330.

Price, L.B., Stegger, M., Hasman, H., Aziz, M., Larsen, J., Andersen, S., Pearson, T., Laurent, F., Keim, P., Skov, R., Aarestrup, F.M., 2012. *Staphylococcus aureus* CC398: host adaptation and emergence of methicillin resistance in livestock. MBio 3 (1), e00305–e00311.

Reinoso, E., Bettera, S., Frigerio, C., Direnzo, M., Calzolari, A., Bogni, C., 2004. RAPD-PCR analysis of *Staphylococcus aureus* strains isolated from bovine and human hosts. Microbiol. Res. 159, 245–255. http://dx.doi.org/10.1016/j.micres.2004.04.002.

Ronco, T., Stegger, M., Pedersen, K., 2017. Draft genome sequence of a sequence type 398 methicillin-resistant *Staphylococcus aureus* isolate from a danish dairy cow with mastitis. Genome Announc. 5, e00492–17.

Sakwinska, O., Giddey, M., Moreillon, M., Morisset, D., Waldvogel, A., Moreillon, P., 2011. *Staphylococcus aureus* host range and human-bovine host shift. Appl. Environ. Microbiol. 77, 5908–5915. http://dx.doi.org/10.1128/AEM.00238-11.

Santos Costa, S., Viveiros, M., Rosato, A.E., Melo-Cristino, J., Couto, I., 2015. Impact of efflux in the development of multidrug resistance phenotypes in *Staphylococcus aureus*. BMC Microbiol. 15, 232–247. http://dx.doi.org/10.1186/s12866-015-0572-8.

Smith, E.M., Green, L.E., Medley, G.F., Bird, H.E., Fox, L.K., Schukken, Y.H., Kruze, J.V., Bradley, A.J., Zadoks, R.N., Dowson, C.G., 2005. Multilocus sequence typing of intercontinental bovine *Staphylococcus aureus* isolates. J. Clin. Microbiol. 43, 4737–4743. http://dx.doi.org/10.1128/JCM.43.9.4737.

Spoor, L.E., Mcadam, P.R., Weinert, L.A., Rambaut, A., Hasman, H., Aarestrup, F.M., Kearns, A.M., Larsen, A.R., Skov, R.L., Ross, J., 2013. Livestock origin for a human pandemic clone of community- associated methicillin-resistant *Staphylococcus aureus*. MBio 4, e00356–13. http://dx.doi.org/10.1128/mBio.00356-13.

Umeda, K., Nakamura, H., Yamamoto, K., Nishina, N., Yasufuku, K., Hirai, Y., Hirayama, T., Goto, K., Hase, A., Ogasawara, J., 2017. Molecular and epidemiological characterization of staphylococcal foodborne outbreak of *Staphylococcus aureus* harboring *seg, sei, sem, sen, seo*, and *selu* genes without production of classical enterotoxi. Int. J. Food Microbiol. 256, 30–35. http://dx.doi.org/10.1016/j.ijfoodmicro.2017.05.023.

Viana, D., Blanco, J., Tormo-más, M.Á., Selva, L., Guinane, C.M., Baselga, R., Corpa, J.M., Lasa, Í., Novick, R.P., Fitzgerald, J.R., Penadés, J.R., 2010. Adaptation of *Staphylococcus aureus* to ruminant and equine hosts involves SaPI-carried variants of von Willebrand factor-binding protein. Mol. Biol. Evol. 77, 1583–1594. http://dx.doi.org/10.1111/j.1365-2958.2010.07312.x.

Xu, J., Tan, X., Zhang, X., Xia, X., Sun, H., 2015. The diversities of staphylococcal species, virulence and antibiotic resistance genes in the subclinical mastitis milk from a single Chinese cow herd. Microb. Pathog. 88, 29–38. http://dx.doi.org/10.1016/j.micpath.2015.08.004.

Yamada, T., Tochimaru, N., Nakasuji, S., Hata, E., 2005. Leukotoxin family genes in *Staphylococcus aureus* isolated from domestic animals and prevalence of *lukM –lukF-PV* genes by bacteriophages in bovine isolates. Vet. Microbiol. 110, 97–103. http://dx.doi.org/10.1016/j.vetmic.2005.07.006.

Yoong, P., Torres, V.J., 2014. The effects of *Staphylococcus aureus* leukotoxins on the host: cell lysis and beyond. Curr. Opin. Microbiol. 16, 63–69. http://dx.doi.org/10.1016/j.mib.2013.01.012.

Zadoks, R.N., Middleton, J.R., McDougall, S., Katholm, J., Schukken, Y.H., 2011. Molecular epidemiology of mastitis pathogens of dairy cattle and comparative relevance to humans. J. Mammary Gland Biol. Neoplasia 16, 357–372. http://dx.doi.org/10.1007/s10911-011-9236-y.

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., Aarestrup, F.M., Larsen, M.V., 2012. Identification of acquired antimicrobial resistance genes. J. Antimicrob. Chemother. 67, 2640–2644. http://dx.doi.org/10.1093/jac/dks261.

TABLE S1 Genes associated with SaPIbov

Gene	Product
sel	staphylococcal enterotoxin-like
sec-bov	staphylococcal enterotoxin C-bovine
orf3	orf3
tst	toxic shock syndrome toxin-1
Orf5	similar to phage terminase small subunit
Orf6	Orf6
Orf7	Orf7
Orf8	Orf8
Orf9	Orf9
Orf10	Orf10
Orf11	Orf11
Orf12	Orf12
Orf13	Orf13
Orf14	Orf14
Orf15	Orf15
Orf16	Orf16
Orf17	Orf17
Orf18	Orf18
Orf19	Orf19
Orf20	Orf20
integrase	integrase-like protein

The table shows 21 genes and their predicted products carried by the 15.9 kbp MGE SaPIBov (acc. no AF217235)

TABLE S2 Genes associated with SaPIbov2

Gene	Product
Or1	Orf1
Orf2	Orf2
Orf3	Orf3
Orf4	similar to ABC transporter ATP-binding protein
Orf5	Orf5
Orf6	Orf6
transposase	transposase
bap	biofilm-associated surface protein
Orf9	Orf9
Orf10	Orf10
Orf11	bacteriophage terminase small subunit
Orf12	Orf12
Orf13	Orf13
Orf14	Orf14
Orf15	Orf15
Orf16	Orf16
Orf17	Orf17
Orf18	Orf18
Orf19	Orf19
Orf20	Orf20
Orf21	transcription regulator
Orf22	Orf22
Orf23	putative phage-associated protein
sip	integrase protein

The table shows 24 genes and their predicted products carried by the 27.0 kbp MGE SaPIbov2 (acc. no AY220730)

TABLE S3 Genes associated with SaPIbov4

Gene	Product
int	integrase
stl	transcription leftward
str	transcription rightward
xis	excisionase
orf5	hypothetical protein
of6	hypothetical protein
pri	primase-like protein
orf8	hypothetical protein
orf9	hypothetical protein
orf10	hypothetical protein
orf11	hypothetical protein
orf12	hypothetical protein
orf13	hypothetical protein
orf14	hypothetical protein
vwb	von Willebrand binding protein
scn	complement inhibitor
orf17	hypothetical protein
orf18	hypothetical protein

The table shows 18 genes and their predicted products carried by the 14.0 kbp MGE SaPIbov4 (acc. no HM211303)

TABLE S4 Genes associated with SaPIbov5

Gene	Product
int	integrase
stl	transcription leftward
str	transcription rightward
xis	excisionase
orf5	hypothetical protein
orf6	hypothetical protein
pri	primase-like protein
orf8	hypothetical protein
orf9	hypothetical protein
orf10	hypothetical protein
orf11	hypothetical protein
orf12	hypothetical protein
ada	adenosin deaminase
scn	complement inhibitor
vwb	von Willebrand binding protein
orf16	hypothetical protein
orf17	hypothetical protein

The table shows 17 genes and their predicted products carried by the 13.5 kbp MGE SaPIbov5 (acc. no HM228919)

TABLE S5 Genes associated with SaPIbov-*υ*Sa*α*

Gene	Product
SAB0342c	probable integrase
SAB0343c	Orf20
SAB0344	Orf19
SAB0345	Orf18
SAB0346	hypothetical protein
SAB0347	hypothetical protein
SAB0348	Orf17
SAB0349	Orf16
SAB0350	Orf15
SAB0351	Orf13 and Orf14
SAB0352	Orf12
SAB0353	Orf11
SAB0354	Orf10
SAB0355	Orf9
SAB0356	Orf8
SAB0357	Orf7
SAB0358	Orf6
SAB0359	Orf15
tst	Toxic shock syndrome toxin-1
SAB0361	hypothetical protein
SAB0362	Orf3
sec-bov	enterotoxin C-bovine
sel	enterotoxin L
SAB0365c	hypothetical protein
SAB0366c	hypothetical protein
SAB0367c	hypothetical protein
SAB0368	hypothetical protein
SAB0369c	hypothetical protein
SAB0370	hypothetical protein
SAB0371	conserved hypothetical protein
SAB0372	conserved hypothetical protein
SAB0373c	conserved hypothetical protein
SAB0374c	hypothetical protein
SAB0375c	hypothetical protein
Set11	exotoxin 11
Set10	exotoxin 10
Set9	exotoxin 9
Set7	exotoxin 7
Set5	exotoxin 5
Set4	exotoxin 4
Set3	exotoxin 3
Set2	exotoxin 2
hsdM	type 1 site-specific deoxyribonuclease
hsdS	type 1 site-specific deoxyribonuclease subunit
Set1	exotoxin 1
SAB0387	probable exported protein
SAB0388c	probable exported protein
SAB0389	probable lipoprotein
SAB0390	probable lipoprotein
SAB0391	hypothetical protein
SAB0392	hypothetical protein
SAB0393	hypothetical protein
SAB0394	hypothetical protein
SAB0395	hypothetical protein
SAB0396	hypothetical protein
SAB0397	hypothetical protein
SAB0398	hypothetical protein

The table shows 57 genes and their predicted products carried by the 45.2 kbp MGE SaPIbov-*υ*Sa*α* which was extracted from strain RF122 (acc. no. AJ938182)

TABLE S6 Genes associated with *φ*SaBov-*υ*Sa*βφ*

Gene	Product
SAB1665c	hypothetical protein
SAB1666c	hypothetical protein
SAB1667c	probable specificity determinant
hsdM	type 1 restriction-modification system M subunit
splF	serine protease
splE	serine protease
splC	serine protease
SAB1672c	serine protease
SAB1673c	serine protease
SAB1674c	hypothetical protein
SAB1675	hypothetical protein
SAB1676c	hypothetical protein
SAB1677c	probable transport protein
SAB1678c	probable transport protein
SAB1679c	probable serine protease precursor
SAB1680c	hypothetical protein
SAB1681c	hypothetical protein
SAB1682c	hypothetical protein
SAB1683c	hypothetical protein
SAB1684	hypothetical protein
SAB1685c	hypothetical protein
lukD	leukotoxcin D subunit
lukE	leukotoxcin E subunit
SAB1688	hypothetical protein
SAB1689c	hypothetical protein
SAB1690	probable membrane protein
SAB1691	probable exported protein
SAB1692	hypothetical protein
SAB1693	hypothetical protein
SAB1694	hypothetical protein
SAB1695	hypothetical protein
seg	enterotoxin G
sen	enterotoxin N
sec-variant	enterotoxin G variant
sei	enterotoxin I
sem-truncated	enterotoxin M
seo	enterotoxin O
SAB1702c	hypothetical protein
SAB1703c	phage-related amidase
SAB1704c	phage-related holin
SAB1705c	phage-related tail fiber
SAB1706c	phage-related cell wall hydrolase
SAB1707c	hypothetical phage-related protein
SAB1708c	hypothetical phage-related protein
SAB1709c	hypothetical phage-related protein
SAB1710c	hypothetical phage-related protein
SAB1711c	hypothetical phage-related protein
SAB1712c	hypothetical phage-related protein
SAB1713c	hypothetical phage-related protein
SAB1714c	hypothetical phage-related protein

TABLE S6 Genes associated with *φ*SaBov-*υ*Sa*βφ* (continued)

Gene	Product
SAB1715c	hypothetical phage-related protein
SAB1716c	hypothetical phage-related protein
SAB1717c	hypothetical phage-related protein
SAB1718c	hypothetical phage-related protein
SAB1719c	hypothetical phage-related protein
SAB1720c	hypothetical phage-related protein
SAB1721c	hypothetical phage-related protein
SAB1722c	phage-related head protein
SAB1723c	hypothetical phage-related protein
SAB1724c	hypothetical phage-related protein
SAB1725c	hypothetical phage-related protein
SAB1726c	phage-associated terminase large subunit
SAB1727c	hypothetical mobile-element-associated protein
SAB1728c	probable mobile-element-associated regulatory protein
rinB	int gene transcriptional activator
SAB1730c	hypothetical phage-related protein
SAB1731c	hypothetical phage-related protein
SAB1732c	hypothetical phage-related protein
SAB1733c	hypothetical phage-related protein
SAB1734c	hypothetical phage-related protein
SAB1735c	hypothetical phage-related protein
SAB1736c	hypothetical phage-related protein
SAB1737c	hypothetical phage-related protein
SAB1738c	hypothetical phage-related protein
SAB1739c	hypothetical phage-related protein
SAB1740c	hypothetical protein
dnaB	replicative DNA helicase
SAB1742c	hypothetical protein
SAB1743c	hypothetical phage-related protein
SAB1744c	hypothetical phage-related protein
SAB1745c	single strand DNA binding protein
SAB1746c	hypothetical phage-related protein
SAB1747c	hypothetical phage-related protein
SAB1748c	hypothetical phage-related protein

TABLE S6 Genes associated with *φ*SaBov-*υ*Sa*βφ* (continued)

Gene	Product
SAB1749c	hypothetical phage-related protein
SAB1750c	hypothetical protein
SAB1751c	hypothetical protein
SAB1752c	hypothetical protein
SAB1753c	phage anti-repressor protein
SAB1754	hypothetical phage-related protein
SAB1755c	phage anti-repressor protein
SAB1756c	hypothetical phage-related protein
SAB1757	probable transcriptional repressor
SAB1758	hypothetical protein
SAB1759	hypothetical protein
SAB1760	integrase

The table shows 96 genes and their predicted products carried by the 80.2 kbp MGE *φ*SaBov-*υ*Sa*βφ* which was extracted from strain RF122 (acc. no. AJ938182)

TABLE S7 Genes associated with *υ*SaBov

Gene	Product
SAB1370c	hypothetical protein
SAB1371c	conserved hypothetical protein
SAB1372c	probable protease
SAB1373c	streptolycin-associated protein sagD homolog
SAB1374c	hypothetical protein
SAB1375c	hypothetical protein
SAB1376c	streptolycin S-associated protein SagB homolog
SAB1377c	hypothetical protein
SAB1378c	hypothetical protein
SAB1379c	probable ABC transporter permease
SAB1380c	probable ABC transport protein
SAB1381c	hypothetical protein
SAB1382c	hypothetical protein
SAB1383c	hypothetical protein

The table shows 14 genes and their predicted products carried by the 7.0 kbp MGE *υ*SaBov which was extracted from strain RF122 (acc. no. AJ938182)

TABLE S8 Genes associated with φ12bov

Gene	Product
SAB0258c	phage-related integrase
SAB0259c	conserved hypothetical protein
SAB0260	hypothetical phage-related protein
SAB0261	hypothetical phage-related protein
SAB0262	hypothetical phage-related protein
SAB0263	hypothetical phage-related protein
SAB0264	hypothetical phage-related protein
SAB0265	type-1 specificity determinant subunit
SAB0266	phage-related holin
SAB0267	phage-related amidase
SAB0268	hypothetical protein
SAB0269	hypothetical protein
SAB0270	hypothetical protein

The table shows 13 genes and their predicted products carried by the 10.8 kbp MGE φ12bov which was extracted from strain RF122 (acc. no. AJ938182)

TABLE S9 Prevalence of virulence genes in isolates from the eight most prevalent sequence types

ST	n	*splA*	*splB*	*splE*	*hlgA*	*hlgB*	*hlgC*	*hlb*	*lukD*	*lukE*	*seu*	*seg*	*sei*	*seo*	*sem*	*sen*	*sel*	*sec*	*tst*	*seh*	*scn*	*sak*	*seq*	*sek*	*sea/sep*	*icaD*	*nuc*	*fib*
151	30	100	97	0	90	100	93	97	97	90	100	100	93	97	93	93	0	0	0	0	0	0	0	0	0	100	100	93
133	14	93	100	0	100	100	100	100	100	100	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	100	100	100
97	10	100	100	20	90	100	90	100	70	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	100	90
479	9	100	100	11	100	100	100	100	100	89	89	78	89	78	89	89	0	0	0	0	0	0	0	0	0	89	100	100
71	8	100	100	88	100	100	100	100	75	75	0	0	13	0	0	0	0	0	0	0	0	0	0	0	0	13	88	100
50	8	100	100	100	88	100	100	100	100	100	88	0	88	75	88	100	0	0	0	0	0	0	0	0	0	100	100	100
3891*	26	96	100	0	96	100	92	100	100	85	96	0	85	92	100	92	0	0	0	0	0	0	0	0	0	100	96	92
3897*	8	100	100	0	100	100	100	100	88	75	100	100	88	100	100	100	0	0	0	0	0	0	0	0	0	100	100	88

The table shows the prevalence (%) of virulence genes identified in isolates from the eight most prevalent STs. The ST97 isolates carried no enterotoxin genes and only 1/7 (7%) of the ST133 isolates and 1/8 (13%) of the ST71 isolates carried a single enterotoxins gene (*sei*). New STs are marked with an asterisk. Only the six most prevalent enterotoxin genes (marked in dark gray) were found among the isolates